

Université d'Évry-Val d'Essonne  
Ecole Doctorale  
des Génomes Aux Organismes



## THÈSE

Numéro 2011EVRY0009

Présentée pour obtenir le grade de :  
**Docteur en sciences de l'université d'Évry-Val d'Essonne**  
Spécialité : biologie moléculaire et cellulaire

**Caractérisation de la variation phénotypique de la  
taille de la baie chez la vigne *Vitis vinifera* L. et  
approches de génétique d'association et de recherche  
de traces de sélection pour ce caractère**

par

**Cléa Houel**

le 29 avril 2011

## JURY

<b>Mathilde Causse</b>	<b>Directeur de Recherche, INRA Avignon</b>	<b>Rapporteur</b>
<b>Christian Chevalier</b>	<b>Directeur de Recherche, INRA Bordeaux</b>	<b>Rapporteur</b>
<b>Nathalie Ollat</b>	<b>Ingénieur de Recherche, INRA Bordeaux</b>	<b>Examineur</b>
<b>Francis Quétier</b>	<b>Professeur, Université d'Évry</b>	<b>Président du jury</b>
<b>Anne-Françoise Adam-Blondon</b>	<b>Directeur de Recherche, INRA Evry</b>	<b>Directeur de thèse</b>





## Remerciements

*Je voudrais tout d'abord remercier Anne-Françoise pour ce passionnant sujet de thèse. Merci de m'avoir fait confiance pour ce travail. Merci pour ta gentillesse, pour tous tes conseils, ta méthodologie qui m'ont permis d'avancer efficacement. Merci pour ton optimisme de tous les jours, pour ne jamais dévaloriser les autres, et toujours nous pousser vers l'avant. Merci pour ces 3 années et demi merveilleuses dans ton équipe.*

*Je voudrais ensuite remercier Bénédicte Sturbois, sans qui cette longue aventure qui est la thèse n'aurait jamais débuté. Tout d'abord pour m'avoir donné le goût de la biologie végétale lors de ces enseignements, puis pour son aide à la préparation du concours pour la bourse MERT.*

*Merci à Aurélie Canaguier, Isabelle Le Clainche, Cécile Guichard et Rémi Bounon et à tous les autres membres de l'équipe pour leurs aides plus que précieuses, sans eux je n'aurais pu obtenir tous les résultats présentés dans cette thèse. J'espère quand même que vous garderez un bon souvenir de mon passage malgré toutes les choses pénibles que vous avez fait pour moi (extraction ADN de bois au mortier, broyage de baies une à une au danguomo, dilution sur mesure de 384 ADN pour le génotypage snplex...). Merci à Cécile pour son efficacité et sa rapidité pour les nombreux scripts Perl qui ont été essentiels dans le traitement de mes données et pour la réalisation de nombreux tableaux, et un grand merci pour la relecture de mon manuscrit. Je remercie également toute l'équipe de l'EPGV pour toutes les belles séquences que j'ai analysé et pour avoir fait passer mes ADN en priorité!*

*Un grand merci à Patrice This, Laurent Torregrosa, Loïc Le Cunff, Thierry Lacombe, Roberto Bacilieri, Stéphane Nicolas, Amidou N'Diaye et à toutes les autres personnes de l'UMR DAVEM de l'INRA de Montpellier pour m'avoir chaleureusement accueillie lors de mes nombreux passages saisonniers et pour les conseils et discussions que nous avons eu sur les résultats et l'avancement de ma thèse. Merci à Geneviève Conéjéro pour m'avoir accueillie sur la plateforme d'imagerie du CIRAD et pour sa gentillesse. Je remercie également toutes les personnes du Domaine expérimental de Vassal et plus particulièrement Sandrine Lalet et Mireille Dessup. Merci à Lisette Lopez pour les délicieux repas qui nous attendaient chaque midi après une longue matinée passée au soleil dans les vignobles...*

*Merci à Nilo Méjià, Marie-Stéphanie Verneret et Jamila Chaïb pour votre collaboration dans mon travail de thèse.*

*Merci à ma grande amie et collègue Imen pour tout ce que l'on a partagé pendant ces 3 années et demi, pour ton amitié, ton soutien, ta folie de tous les jours qui mettait de la joie dans le labo, pour toutes nos discussions non scientifiques le midi qui me permettaient de me déconnecter un peu, et pour tout le reste...*

*Merci également à tous mes autres collègues : Laure pour ces délicieux gateaux et macarons et pour nos nombreuses courses à pied ;- ) Nathalie pour toutes ses réponses aux questions du jour, pour son aide à la préparation des TP et TD, Harry pour avoir été le fournisseur officiel de thé de ma thèse et pour toutes ces petites attentions, Jérémy pour toutes ces délicieuses expériences culinaires, Mathieu pour être un faux « ours mal léché », Ouda pour tes conseils lors de la rédaction de ma thèse, Cécile et Fred pour les petites astuces sur word et excel qui m'ont fait gagner un temps fou. Merci également à Souha, Smahane, Cyrile et Clément et bon courage pour vos thèses à venir. Merci à Mélanie, Odile, Samuel, Karima, Fatima, Samira, Aloïs, Marie-laure, Richard, Philippe, Patricia, Boulos, Arnaud et à toutes les autres personnes de l'URGV que je n'ai pas cité mais qui ne sont pas moins importantes.*

*Je remercie tous mes collègues de l'université d'Evry Val d'Essonne et plus particulièrement Julie Gombert et Isabelle Baurès avec qui ça a été un plaisir de travailler.*

*Merci à mes copines, Laure pour la relecture de mon manuscrit et Laura pour notre soutien téléphonique hebdomadaire et bon courage pour ta thèse à venir...*

*Bien évidemment je remercie ma mère sans qui je n'aurais pu arriver là, mon frère pour son aide inestimable pour les références bibliographiques, ainsi que ma petite belle-sœur, pour m'avoir soutenu les derniers mois de ma thèse qui n'ont pas été faciles.*

*Je dédie cette thèse à mon père, Patrick Houel ♥*

# Table des matières

Remerciements .....	2
Liste des figures .....	7
Liste des tableaux .....	10
Sigles et abréviations.....	11
Avant-propos .....	13
<b>1. Introduction .....</b>	<b>14</b>
<b>1.1. <i>Vitis vinifera</i> L.....</b>	<b>15</b>
1.1.1. Origine et taxonomie .....	15
1.1.2. Variabilité génétique et structuration des populations chez la vigne .....	16
1.1.2.1. Diversité génétique au sein du genre <i>Vitis</i> .....	16
1.1.2.1.1. Diversité génétique estimée à partir de marqueurs SSR .....	16
1.1.2.1.2. Diversité génétique estimée à partir de marqueurs SNP.....	17
1.1.2.1.3. Séquençage du génome de la vigne .....	18
1.1.2.2. Structure des populations de <i>Vitis vinifera</i> L.....	19
1.1.3. Domestication et sélection de la vigne .....	22
1.1.3.1. De la domestication à la viticulture moderne: petit historique .....	22
1.1.3.2. La domestication.....	22
1.1.3.3. La sélection.....	25
<b>1.2. Analyse du déterminisme génétique et moléculaire de la variabilité d'un caractère .....</b>	<b>28</b>
1.2.1. Détection de QTL dans des populations issues de croisements contrôlés.....	29
1.2.2. Génétique d'association.....	31
1.2.2.1. Déséquilibre de liaison et test d'association : généralités.....	31
1.2.2.2. Déséquilibre de liaison et test d'association chez la vigne .....	33
1.2.3. Recherche de traces de sélection .....	35
1.2.3.1. Généralités .....	35
1.2.3.2. Exemples d'identification de traces de sélection.....	36
<b>1.3. Bases physiologiques et génétiques de la variabilité des fruits charnus.....</b>	<b>39</b>
1.3.1. Développement floral chez la vigne .....	39
1.3.1.1. Structure des fleurs de vigne.....	39
1.3.1.2. Bases moléculaires du développement de la fleur de vigne .....	40
1.3.2. Développement des fruits charnus.....	41
1.3.2.1. Anatomie de la baie de vigne et de tomate .....	42
1.3.2.2. Développement de la baie de vigne .....	44
1.3.2.3. Développement de la baie de tomate .....	45
1.3.3. Etude des déterminants de la variabilité de la taille des fruits charnus .....	47
1.3.3.1. Variation de la taille de la baie chez la tomate .....	48
1.3.3.1.1. Déterminants physiologiques.....	48
1.3.3.1.2. Déterminants génétiques et moléculaires.....	50
1.3.3.2. Variation de la taille de la baie chez la vigne .....	52
1.3.3.2.1. Déterminants physiologiques.....	52
1.3.3.2.2. Déterminants génétiques et moléculaires.....	55
<b>1.4. Contexte et objectifs de la thèse .....</b>	<b>58</b>
<b>2. Matériel et méthodes .....</b>	<b>60</b>
<b>2.1. Matériel végétal .....</b>	<b>61</b>
<b>2.2. Méthodes .....</b>	<b>63</b>
2.2.1. Phénotypage pour l'étude des déterminants de la variation de la taille de la baie .....	63
2.2.1.1. Analyse macroscopique des ovaires et des baies de vigne .....	63
2.2.1.1. Analyse microscopique des ovaires et des baies de vigne.....	64
2.2.1.1.1. Préparation des échantillons.....	64
2.2.1.1.2. Analyse microscopique.....	65
2.2.2. Préparation du matériel pour les expériences d'hybridation fluorescente <i>in situ</i> de clones BAC (BAC-FISH) .....	66
2.2.2.1. Choix du matériel végétal .....	66
2.2.2.2. Choix des sondes .....	67
2.2.3. Extraction et quantification d'ADN génomique.....	68

2.2.3.1. A partir de feuilles .....	68
2.2.3.2. A partir de bois .....	68
2.2.3.3. A partir de baies .....	69
2.2.3.4. Quantification de l'ADN et vérification de sa qualité .....	70
2.2.4. Extraction d'ARN et transcription inverse .....	70
2.2.4.1. Extraction ARN .....	70
2.2.3.2. Quantification et transcription inverse .....	71
2.2.5. PCR .....	71
2.2.5.1. Définition des couples d'amorces .....	71
2.2.5.2. Test génomique .....	73
2.2.5.3. Séquençage .....	73
2.2.5.4. PCR quantitative .....	74
2.2.5.4.1. Sur ADN génomique .....	74
2.2.5.4.2. Sur ADN complémentaire .....	75
2.2.6. Analyse du polymorphisme de séquence .....	76
2.2.6.1. Détection du polymorphisme de séquence .....	76
2.2.6.2. Analyse de la diversité nucléotidique, du déséquilibre de liaison et test de neutralité .....	76
2.2.6.3. Choix des SNP .....	77
2.2.7. Génotypage .....	78
2.2.8. Analyses statistiques et tests d'association .....	78
2.2.8.1. Tests d'association .....	78
2.2.8.2. Analyses statistiques .....	79
<b>3. Analyse phénotypique fine de la variation de la taille de la baie chez <i>V. vinifera</i> .....</b>	<b>80</b>
3.1. Introduction .....	81
3.2. Définition d'une stratégie pour l'estimation du nombre de cellules de la baie .....	82
3.3. Description de la variabilité de la taille de la baie chez <i>V. vinifera</i> .....	84
Article: <i>Genetic variability of the berry size in grapevine (Vitis vinifera L.) involves ovary and early berry development associated factors</i> .....	85
3.4. Conclusion du chapitre .....	103
<b>4. Recherche de gènes impliqués dans la variation de la taille de la baie chez <i>V. vinifera</i> ssp <i>sativa</i>: étude de génétique d'association .....</b>	<b>105</b>
4.1. Introduction .....	106
4.2. Analyse génétique de la variation de la taille des baies .....	106
Article: <i>Screening candidate genes for berry traits using association genetics in grapevine (Vitis vinifera L.)</i> .....	107
<b>5. Recherche de traces de domestication en relation avec la variation de la taille de la baie chez <i>V. vinifera</i>: étude du locus <i>Fleshless berry</i> .....</b>	<b>129</b>
5.1. Introduction .....	130
5.2. Analyse de la diversité génétique au locus <i>Flb</i> .....	131
Article: <i>Patterns of sequence polymorphism in the Fleshless berry locus in cultivated and wild Vitis vinifera L. accessions</i> .....	132
5.3. Analyses complémentaires .....	154
5.3.1. Diversité génétique au locus <i>Flb</i> chez <i>V. vinifera</i> ssp <i>sylvestris</i> .....	154
5.3.2. Génétique d'association dans les 3 sous-populations .....	156
5.3.3. Expression des gènes candidats au cours du développement de l'ovaire et de la baie .....	157
5.4. Conclusion du chapitre .....	158
<b>6. Discussion générale et perspectives .....</b>	<b>161</b>
<b>7. Références bibliographiques .....</b>	<b>172</b>
<b>8. Annexes .....</b>	<b>198</b>
Annexe 1: Figures et tableaux supplémentaires de l'article: <i>Genetic variability of the berry size in grapevine (Vitis vinifera L.) involves ovary and early berry development associated factors</i> (Chapitre 3) .....	199

<b>Annexe 2:</b> Figures et tableaux supplémentaires de l'article: <i>Screening candidate genes for berry traits using association genetics in grapevine (Vitis vinifera L.)</i> (Chapitre 4) .....	<b>207</b>
<b>Annexe 3:</b> Figures et tableaux supplémentaires de l'article: <i>Patterns of sequence polymorphism in the Fleshless berry locus in cultivated and wild Vitis vinifera accessions</i> (Chapitre 5).....	<b>219</b>
<b>Annexe 4:</b> Figures et tableaux des résultats supplémentaires réalisés sur la région <i>Fleshless berry</i> (Chapitre 5) .....	<b>234</b>



# Liste des figures

<b>Figure 1.1:</b> Taxonomie simplifiée des eudicotylédones.....	15
<b>Figure 1.2:</b> Taxonomie simplifiée de <i>Vitis vinifera</i> L.....	15
<b>Figure 1.3:</b> Utilisation des données de diversité génétique et de structure des populations de <i>V. Vinifera</i> pour définir des sous-ensembles de génotypes pour différentes applications.....	19
<b>Figure 1.4:</b> Fresque décrivant les vendanges, Thèbes, tombe de Nakht.....	22
<b>Figure 1.5:</b> Diffusion de la vigne cultivée d'après Fregoni (1991).....	22
<b>Figure 1.6:</b> Principales différences morphologiques entre la vigne sauvage et cultivée au niveau des feuilles, fleurs, grappes à maturité et pépins.....	23
<b>Figure 1.7:</b> Schéma simplifié des étapes de domestication et de sélection de la vigne.....	25
<b>Figure 1.8:</b> Relations de premier degré d'apparentement entre 583 génotypes <i>V. vinifera</i> cultivés uniques provenant de la collection de l'USDA.....	25
<b>Figure 1.9:</b> Variation clonale dans la famille des Pinot.....	26
<b>Figure 1.10:</b> Comparaison des populations utilisées en cartographie de QTL et dans des études de génétique d'association.....	30
<b>Figure 1.11:</b> Représentation schématique du déséquilibre de liaison et de l'équilibre de liaison entre deux loci.....	31
<b>Figure 1.12:</b> Relation entre l'étendu du DL et la résolution d'études de génétique d'association.....	32
<b>Figure 1.13:</b> Effet de la domestication sur la diversité génétique des plantes.....	35
<b>Figure 1.14:</b> Différentes stratégies utilisées pour l'étude du déterminisme génétique de caractères quantitatifs .....	36
<b>Figure 1.15:</b> Exemples de traces de sélection dans la région du gène <i>tb1</i> chez le maïs et du locus <i>waxy</i> chez le riz.....	37
<b>Figure 1.16:</b> La fleur de vigne.....	39
<b>Figure 1.17:</b> Tissus de l'ovaire.....	39
<b>Figure 1.18:</b> Modèle ABCDE indiquant les différents domaines d'activité des gènes des classes A, B, C D et E dans le développement des pièces florales chez <i>Arabidopsis</i> .....	40
<b>Figure 1.19:</b> Classification simplifiée des fruits.....	42

<b>Figure 1.20:</b> Anatomie de la baie de vigne et de la tomate.....	42
<b>Figure 1.21:</b> Photo d’une grappe de raisins contenant des baies millerandées distinguables par leur petite taille.....	44
<b>Figure 1.22:</b> Différents stades du développement de la baie de vigne.....	45
<b>Figure 1.23:</b> Différentes phases du développement de la baie de tomate.....	46
<b>Figure 1.24:</b> Variation de la taille du fruit chez la tomate.....	48
<b>Figure 1.25:</b> Illustration de la gamme de variation de la taille et de la forme de la baie chez la vigne cultivée.....	52
<b>Figure 1.26:</b> Morphologie du mutant <i>Fleshless berry</i> .....	53
<b>Figure 1.27:</b> Organisation chimérique du mutant <i>Fleshless berry</i> .....	56
<b>Figure 2.1:</b> Matériel végétal utilisé au cours de ma thèse.....	61
<b>Figure 2.2:</b> Stades du développement ciblés pour le phénotypage macroscopique des ovaires et des baies.....	63
<b>Figure 2.3:</b> Phénotypage des ovaires et du diamètre des baies.....	64
<b>Figure 2.4:</b> Estimation du nombre et du volume des cellules des baies à véraison.....	65
<b>Figure 2.5:</b> Matériel végétal récolté pour les expériences de BAC-FISH.....	66
<b>Figure 2.6:</b> Stades de développement auxquels ont été faits les prélèvements d’ovaires et de baies pour les extractions d’ARN.....	70
<b>Figure 2.7:</b> Exemple d’alignement des séquences et de détection des SNP dans le sous-échantillon SNPGrpMap à l’aide du logiciel Genalys®.....	76
<b>Figure 3.1:</b> Localisation des différents sites dans lesquels les ovaires et les baies ont été prélevés et préparés pour l’ensemble des mesures réalisées.....	82
<b>Figure 3.2:</b> Corrélation entre le nombre moyen de cellules comptées sur des coupes transversales de baies observées au microscope optique et la quantité d’ADN extraite par baie dans un échantillon de 26 génotypes cultivés.....	83
<b>Figure 3.3:</b> Photo d’une vigne ravagée par les lapins.....	104
<b>Figure 5.1:</b> Différences morphologiques de la baie et du pépin entre l’Ugni Blanc et l’Ugni Blanc mutant et la vigne cultivée et sauvage.....	130
<b>Figure 5.2:</b> Répartition et fréquence allélique des SNP génotypés dans les différents échantillons.....	155



<b>Figure 5.3:</b> Test d'association pour la variation du poids de la baie.....	156
<b>Figure 5.4:</b> Quantité relative de transcrits correspondant aux gènes VVC2897A, VV05785A et VV05786A, normalisée sur l'expression des gènes de référence au cours du développement de la fleur et de la baie, moyennée sur deux années.....	157
<b>Figure 6.1:</b> Schéma explicatif des paramètres de l'ovaire et de la baie à l'origine de la variation du poids de la baie chez la vigne <i>Vitis vinifera</i> L.....	162
<b>Figure 6.2:</b> Schéma hypothétique de la domestication de la vigne <i>Vitis vinifera</i> L. ayant porté sur la taille des baies en fonction de leur utilisation et de leur origine géographique, et cycles de sélection pour la taille de la baie.....	168

## Liste des tableaux

<b>Tableau 1.1:</b> Etudes les plus significatives de la diversité génétique des compartiments sauvages et cultivés de <i>V. vinifera</i> et d'autres espèces appartenant à la famille des Vitacées à l'aide de marqueurs SSR.....	16
<b>Tableau 1.2:</b> Etudes les plus significatives de la diversité génétique des compartiments sauvages et cultivés de <i>V. vinifera</i> et d'autres espèces appartenant à la famille des Vitacées à partir de marqueurs SNP.....	17
<b>Tableau 1.3:</b> Liste des mutations somatiques permettant la distinction morphologique entre clones d'une même variété dont le gène muté a été identifié.....	27
<b>Tableau 1.4:</b> Détections de QTL réalisées chez la vigne pour des caractères d'intérêts agronomiques.....	30
<b>Tableau 1.5:</b> Liste des orthologues putatifs des gènes de classe A, B, C, D et E chez la vigne.....	40
<b>Tableau 1.6:</b> Comparaison entre les différents stades du développement de la baie de vigne et de tomate.....	47
<b>Tableau 1.7:</b> Ensemble des études QTL pour le caractère poids de la baie chez la vigne.....	55
<b>Tableau 5.1:</b> Critères à l'origine du choix des trois gènes pour l'étude transcriptomique...	156

## Sigles et abréviations

ADN	: Acide DésoxyRibonucléique
ADNc	: ADN complémentaire
ADNg	: ADN génomique
ARN	: Acide RiboNucléique
ARNm	: ARN messenger
BAC	: ( <i>Bacterial Artificial Chromosome</i> ) chromosome artificiel bactérien
cm	: centimètre
cM	: centiMorgan
CE	: variété de raisin de Cuve originaire d'Europe de l'Est
CO	: variété de raisin de Cuve originaire d'Europe de l'Ouest
Cv	: Cultivar
DL	: ( <i>Disequilibrium Linkage</i> ) déséquilibre de liaison
dNTP	: desoxy-Nucleotide-TriPhosphates
DTT	: DiThioThréitol
EDTA	: ( <i>Ethylene Diamine Tetraacetic Acid</i> ) acide tétraacétique diamine éthylène
EST	: ( <i>Expressed Sequenced Tag</i> ) étiquettes de transcrits
FISH	: ( <i>Fluorescence In Situ Hybridation</i> ) hybridation fluorescente <i>in situ</i>
<i>Flb</i>	: <i>Fleshless berry</i>
g	: gramme
GA	: ( <i>Gibberellic Acid</i> ) acide gibbérellique
GLM	: ( <i>Generalized Linear Model</i> ) modèle linéaire généralisé
GS	: ( <i>Genomic Selection</i> ) recherche de traces de sélection sur le génome
h	: heure
HMGR	: 3-Hydroxy-3-Methylglutaryl-CoA Reductase
IGGP	: <i>International Grape Genome Program</i>
INDEL	: INsertion ou DELétion
IFVV	: Institut Français de la Vigne et du Vin
JC	: Jésus Christ
kb	: 10 <sup>3</sup> paires de bases
L	: Litre
M	: Mole
Mb	: 10 <sup>6</sup> paires de bases

mg	: milligramme
min	: minute
ml	: millilitre
MLM	: ( <i>Mixed Linear Model</i> ) modèle linéaire mixte
mM	: milliMolaire
mm	: millimètre
NAM	: ( <i>Nested Association Mapping</i> ) génétique d'association emboîtée
ng	: nanogramme
OIV	: Organisation Internationale de la Vigne et du Vin
oligodT	: oligodeoxyThymidylique
pb	: paire de bases
PCR	: ( <i>Polymerisation Chain Reaction</i> ) réaction en chaîne de polymérisation
pmol	: picomole
q-PCR	: ( <i>quantitative PCR</i> ): PCR quantitative
q-RT-PCR	: ( <i>quantitative Real Time PCR</i> ) PCR quantitative en temps réel
QTL	: ( <i>Quantitative Trait Locus</i> ) région ou gène contrôlant un caractère quantitatif
RIL	: ( <i>Recombinant Inbred Line</i> ) lignée recombinante
RT	: ( <i>Reverse Transcription</i> ) Transcription inverse
s	: seconde
SNP	: ( <i>Single Nucleotide Polymorphism</i> ) mutation ponctuelle
SSR	: ( <i>Short sequence repeate</i> ) petite séquence répétée ou microsatellite
TE	: ( <i>Transposable Element</i> ) élément transposable
TE	: variété de raisin de Table originaire d'Europe de l'Est
Tm	: ( <i>melting Temperature</i> ) température d'hybridation
UICN	: Union Internationale pour la Conservation de la Nature
USDA	: <i>United States Department of Agriculture</i>
UTR	: ( <i>UnTranslated Region</i> ) région non-traduite
<i>V. vinifera</i>	: <i>Vitis vinifera</i> L.
WGAS	: ( <i>Whole Genome Association Scan</i> ) étude de génétique d'association tout génome
µg	: microgramme
µm	: micromètre

## Avant-propos

La vigne (*Vitis vinifera* L.) est l'une des plus importantes espèces fruitières cultivées dans le monde. Avec une production totale de raisins estimée en 2008 à 67 millions de tonnes, dont 71% ont été utilisées pour la production de vin, 27% pour la consommation en frais et 2% pour la production de raisins secs, selon l'Organisation Internationale de la Vigne et du Vin (<http://www.oiv.int>). La superficie des vignobles à travers le monde est estimée à 7,7 millions d'hectares répartis entre l'Europe (58,4%), l'Asie (21,1%), l'Amérique (12,8%), l'Afrique (5,0%) et l'Océanie (2,7%). La France était le deuxième pays producteur de vin dans le monde en 2008 après l'Italie avec une production de 41,4 millions d'hectolitres.

Un des atouts majeurs des vins français est leur typicité et leur qualité. La maîtrise de la qualité de la vendange est donc une nécessité absolue. La taille de la baie est un des facteurs essentiels de la qualité de la vendange. En effet, les meilleurs cultivars de raisins de cuve sont ceux produisant des baies de petites tailles, dû principalement à une dilution plus ou moins importante des composés de la pulpe et de la pellicule. Une large partie du déterminisme de la taille des baies est de nature génétique. Cependant aucun gène impliqué dans la variation de la taille de la baie chez la vigne n'a encore été identifié à ce jour. Les connaissances actuelles des mécanismes génétiques impliqués dans la variation de la taille de la baie chez la vigne sont largement insuffisantes pour manipuler efficacement ce caractère dans des programmes de sélection. Or, la filière viticole est aujourd'hui confrontée à deux défis majeurs: les changements climatiques qui remettent en cause l'association cépage / terroir à la base du développement de la viticulture française et européenne et la demande sociétale de diminution de l'utilisation de produits phytosanitaires en viticulture. Afin de répondre aux contraintes abiotiques et biotiques sans augmenter les coûts de production, la sélection de nouveaux génotypes semble être une solution durable. Ces génotypes devront contenir des résistances durables aux maladies, très généralement introgressées à partir d'espèces américaines et asiatiques et devront être d'une qualité irréprochable. L'amélioration des connaissances du caractère taille de la baie participerait au développement des méthodes de sélection assistée par marqueurs pour une sélection rapide des génotypes les plus intéressants et produisant des baies de grande qualité.

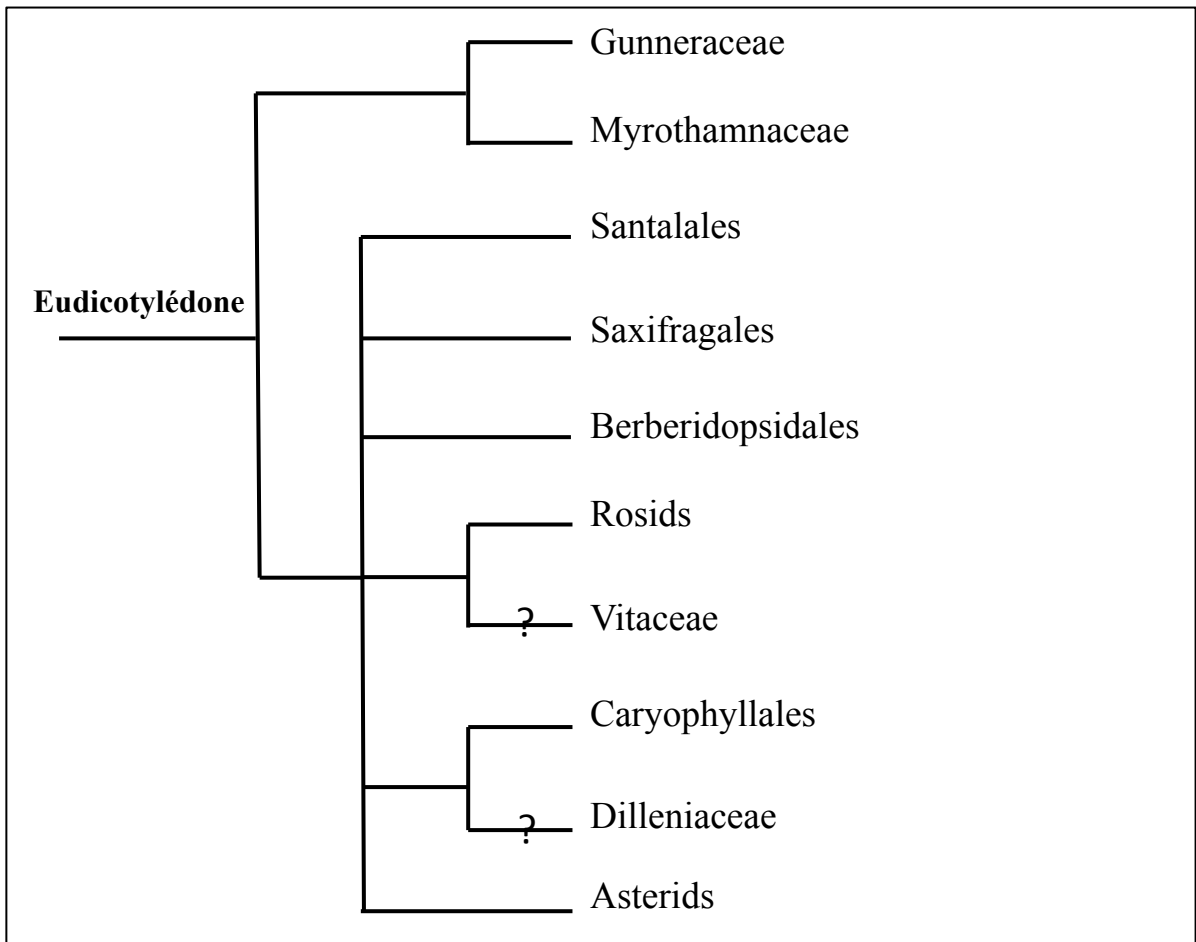
L'objectif de ma thèse était de contribuer à l'amélioration des connaissances sur la variabilité de la taille de la baie chez la vigne, à travers l'identification de déterminants morphologiques et génétiques.



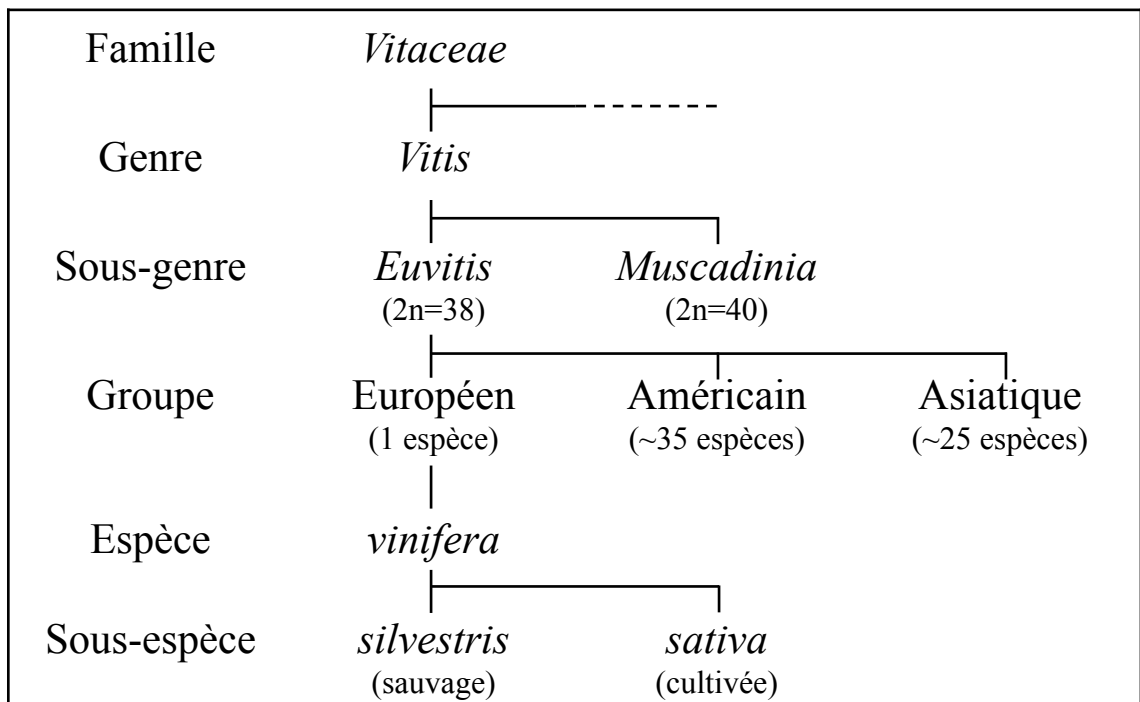
# 1. Introduction



*Photographie d'une inflorescence à l'anthèse.*



**Figure 1.1:** Taxonomie simplifiée des eudicotylédones (source: tolweb.org/).



**Figure 1.2:** Taxonomie simplifiée de *Vitis vinifera* L.



## 1.1. *Vitis vinifera* L.

### 1.1.1. Origine et taxonomie

La vigne (*Vitis vinifera* L.) est une plante pérenne qui se présente sous la forme d'une liane. Elle appartient aux eudicotylédones et à la famille des Vitacées. La position des Vitacées parmi les eudicotylédones est controversée, cette famille appartenant soit à un sous-branchement des rosidés comprenant également le peuplier (Figure 1.1; Savolainen et al, 2000 ; Hilu et al, 2003), soit à une branche basale des core-eudicotylédones. La vigne appartient au genre *Vitis*, le seul genre de cette famille à produire des fruits consommables. Au sein du genre *Vitis*, on distingue deux sous-genres: *Muscadinia* et *Euvitis*, distinction basée principalement sur des différences morphologiques et sur le nombre de chromosomes ( $2n=40$  pour *Muscadinia* et  $2n=38$  pour *Euvitis*; Bouquet, 1982; Galet, 1993). Le sous-genre *Euvitis* auquel appartient *Vitis vinifera* L. (*V. vinifera*) comprend une soixantaine d'espèces réparties selon leur origine géographique entre:

- les espèces d'origine américaine dont certaines sont utilisées directement comme porte-greffe ou dans la constitution d'hybrides résistants au phylloxéra ou bien au mildiou ou à l'oïdium (sources de résistances pour des hybrides producteurs directs ou des variétés améliorées pour la résistance aux maladies; Levadoux, 1956; Pouget, 1990; This et al, 2006);
- les espèces d'origine asiatique dont seul *Vitis amurensis* a été utilisé jusqu'à présent dans des programmes de sélection pour sa résistance au froid ou aux maladies (Olmo, 1976; This et al, 2006; C. Schneider, communication personnelle);
- la seule espèce d'origine européenne (Figure 1.2), *V. vinifera* regroupant l'ensemble des cépages cultivés (*V. vinifera* ssp. *sativa*) et des vignes sauvages aussi appelées lambrusques (*V. vinifera* ssp. *silvestris*).

Le compartiment cultivé de l'espèce *V. vinifera* compte entre 6000 et 10000 variétés (Galet, 2000) qui se distinguent par leurs différences morphologiques (dits caractères ampélographiques) telles que la forme des feuilles, des grappes ou des baies (Alleweldt et Dettweiler, 1994; Boursiquot, 1996). Ces variétés sont utilisées soit pour la vinification, on parle alors de cépages de cuve, soit pour la consommation du fruit frais ou sec et on parle alors de cépages de table. Les fruits de la vigne sont également utilisés de façon plus marginale pour la fabrication de jus de raisin, de confiture, de sirop, d'eau de vie, de vinaigre, d'huile de pépins de raisin, d'acide tartrique mais également dans les cosmétiques ou comme

**Tableau 1.1:** Etudes les plus significatives de la diversité génétique des compartiments sauvages et cultivés de *V. vinifera* et d’autres espèces appartenant à la famille des Vitacées à l’aide de marqueurs SSR.

Compartiment	Nombre de marqueurs	Nombre de génotypes uniques	Nombre d'allèles	Hétérozygotie observée	Référence
Cultivé	4	77	30	—	Bower et al, 1996
Cultivé	9	160	88	0.82	Sefc et al, 2000
Cultivé	8	266	94	0.76	Aradhya et al, 2003
Cultivé	14	58	129	0.66	Persy, 2007
Cultivé	20	2323	338	0.76	Le Cunff et al, 2008; Laucou et al, 2011
Cultivé	34	745	207	0.40	Cipriani et al, 2010
Sauvage (Tunisie)	9	109	66	0.72	Snoussi et al, 2004
Sauvage (France)	—	101	65	0.58	Arrigo et Arnold, 2007
Sauvage	6	310	—	0.67	Grassi et al, 2008
Sauvage	20	72	203	0.62	Laucou et al, 2011
Espèce <i>Vitis</i>	20	72	405	0.78	

engrais.

## 1.1.2. Variabilité génétique et structuration des populations chez la vigne

### 1.1.2.1. Diversité génétique au sein du genre *Vitis*

#### 1.1.2.1.1. Diversité génétique estimée à partir de marqueurs SSR

De très nombreuses études ont été menées afin d'étudier la diversité génétique de la vigne en utilisant des marqueurs microsatellites (*Short Sequence Repeat*: SSR), tant au sein du compartiment cultivé et sauvage que chez des espèces *Vitis* non *vinifera* (Bower et al, 1996; Sefc et al, 2000; Aradhya et al, 2003; Snoussi et al, 2004; Persly, 2007; Arrigo et Arnold, 2007; Le Cunff et al, 2008; Grassi et al, 2008; Cipriani et al, 2010; Laucou et al, 2011). Le tableau 1.1 résume les principales données obtenues dans des études utilisant un échantillon de vigne cultivée appartenant à diverses régions géographiques, un échantillon de vigne sauvage et/ou un échantillon d'espèce *Vitis*.

Plusieurs milliers de génotypes de *V. vinifera* ont été analysés, mais le nombre de génotypes et de marqueurs est très différent selon les études. Cependant, on peut noter que toutes les études indiquent un taux d'hétérozygotie très élevé que ce soit chez la vigne cultivée (0.40 à 0.82), chez la vigne sauvage (0.58 à 0.72) ou au niveau inter-spécifique dans le genre *Vitis* (0.78). L'analyse la plus importante qui présente les résultats de 20 marqueurs SSR génotypés sur un échantillon de 2323 génotypes maximisant l'aire géographique de la vigne cultivée (Le Cunff et al, 2008 ; Laucou et al, 2011) a révélé un taux d'hétérozygotie de 0.76 et un nombre moyen d'allèles par locus de 16.9 chez la vigne cultivée. Cette diversité est comparable voir légèrement plus faible à celle de populations sauvages de riz (taux d'hétérozygotie = 0.86; Gao et al, 2006) ou cultivées de maïs (14.8 allèles par locus, taux d'hétérozygotie = 0.79; Vigouroux et al, 2005). Ce taux est néanmoins plus élevé que celui du blé (taux d'hétérozygotie = 0.60; Peng et al, 2009) et de la tomate (taux d'hétérozygotie = 0.49; Ranc et al, 2008). Ces résultats peuvent s'expliquer par le fait qu'à l'état sauvage *V. vinifera* est dioïque et présente de ce fait un niveau élevé d'hétérozygotie liée à une allogamie stricte. Le niveau également élevé d'hétérozygotie dans le compartiment cultivé, procédant des fleurs hermaphrodites entraînant préférentiellement l'autogamie, peut certainement s'expliquer par une contre-sélection des individus les plus homozygotes, certainement moins vigoureux et moins fertiles (Bouquet, 2011). Par ailleurs, la découverte de la multiplication

**Tableau 1.2:** Etudes les plus significatives de la diversité génétique des compartiments sauvages et cultivés de *V. vinifera* et d’autres espèces appartenant à la famille des Vitacées à partir de marqueurs SNP. Pour une des études (Lijavetzky et al, 2007) la diversité génétique a été calculée sur l’ensemble de l’échantillon.

Echantillon	Nombre de gènes	SNP/ nombre de nucléotides séquencés	Diversité nucléotidique (II)	Référence
7 cultivés	25	1/116	0.0084	Salmaso et al, 2004
1 espèce <i>Vitis</i>		1/92	0.0102	
1 hybride interspécifique		1/74	0.0128	
9 cultivés 2 sauvages	230	1/64	0.0051	Lijavetzky et al, 2007
92 cultivés	3	1/49	—	Le Cunff et al, 2007
24 cultivés	35	1/104	—	Vezzulli et al, 2008
10 sauvages	35	1/129		
39 espèces <i>Vitis</i>	35	1/97		

végétative a facilité la pérennisation et la diffusion des génotypes ayant les structures génétiques les plus performantes et a permis l'accumulation de mutations somatiques, contribuant aussi probablement au taux élevé d'hétérozygotie dans le compartiment cultivé.

#### 1.1.2.1.2. Diversité génétique estimée à partir de marqueurs SNP

Contrairement aux marqueurs SSR qui sont considérés généralement comme neutres, les marqueurs basés sur le polymorphisme d'une base (*Single Nucleotide Polymorphism*: SNP) ont tout d'abord été identifiés dans des gènes et sont donc susceptibles d'être soumis à la sélection (Lijavetzky et al, 2006; This et al, 2007). Jusqu'à récemment, leur détection était réalisée par re-séquençage Sanger de fragments de gènes, ce qui reste une technique assez lourde. Le polymorphisme de séquences de fragments d'Acide DésoxyRibonucléique (ADN) a donc été étudié sur un nombre restreint de cultivars (neuf génotypes pour Salmaso et al, 2004; 11 génotypes pour Lijavetzky et al, 2007 et 91 génotypes pour Le Cunff et al, 2008) et pour un nombre relativement restreint de fragments de gènes (25 fragments dans Salmaso et al 2004, 230 fragments dans Lijavetzky et al, 2007 et trois gènes chez Le Cunff et al, 2008). Malgré une grande disparité dans le nombre de fragments séquencés et dans le nombre et la composition des génotypes analysés, l'ensemble des études montre que la diversité génétique chez *V. vinifera* est élevée dans les gènes (régions non-traduites (*Untranslated Region*: UTR, introns et exons confondus)), aussi bien dans le compartiment cultivé que sauvage, avec 1 SNP toutes les 118 bases (Salmaso et al, 2004), 1 SNP toutes les 64 bases (Lijavetzky et al, 2007) et jusqu'à 1 SNP toutes les 49 bases (Le Cunff et al, 2008). La diversité génétique semble cependant être moins élevée dans le compartiment sauvage (Tableau 1.2) avec 1 SNP toutes les 129 bases (Vezzulli et al, 2008) comparé à 1 SNP toutes les 104 bases chez les cultivées pour les mêmes fragments analysés. Comme observé avec les marqueurs SSR, la diversité nucléotidique est encore plus élevée lorsqu'on se place au niveau interspécifique (Tableau 1.2). En effet, le nombre de SNP varie de 1 SNP toutes les 74 bases (Salmaso et al, 2004) et 1 SNP toutes les 97 bases (Vezzulli et al, 2008) quand on considère plusieurs espèces *Vitis*.

Contrairement aux autres plantes domestiquées comme par exemple le maïs et la tomate (Wang et al, 1999; Ranc et al, 2008), *Vitis vinifera* semble présenter une variabilité génétique plus faible dans le compartiment sauvage que dans le compartiment cultivé (Aradhya et al, 2003; Arroyo-Garcia et al, 2006; Grassi et al, 2008). Cette différence de diversité peut



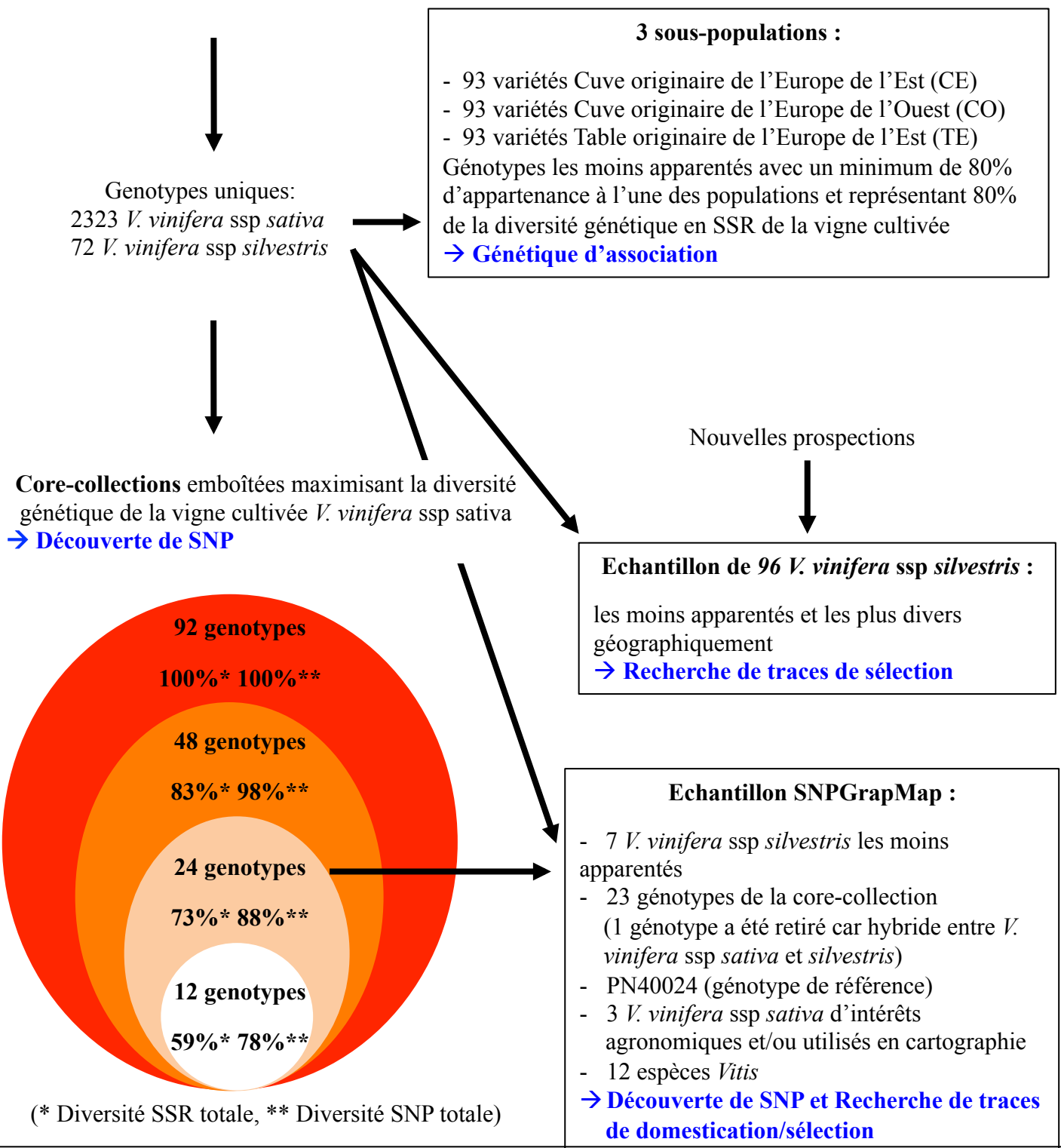
s'expliquer par la différence du nombre de génotypes analysés. En effet, dans toutes les études (Tableau 1.2), le nombre de génotypes sauvages est toujours inférieur à celui des génotypes cultivés et ne représente probablement pas la diversité génétique de ce compartiment. Mais, cette diminution de la diversité dans le compartiment sauvage, pourrait également être liée à l'introduction des agents pathogènes en Europe (*Plasmopara viticola*, *Erysiphe necator* et *Phylloxera vastatrix*), qui ont détruit une grande partie des vignobles européens à la fin du XIX<sup>ème</sup> siècle, et également, une partie des populations de vignes sauvages. La diversité de *V. vinifera* ssp *silvestris* n'est très certainement aujourd'hui que le pâle reflet de ce qui existait avant l'introduction de ces agents pathogènes (Pouget, 1990 ; Arnold et al, 1998). De plus, la petite taille du compartiment sauvage est devenu en lui-même un facteur de vulnérabilité, augmentant l'effet du transfert de gènes venant du compartiment cultivé *via* le pollen (Ellstrand, 2003), même si cet événement est probablement minimisé par une date de floraison différente entre les deux compartiments (This et al, 2006).

En revanche, la diversité génétique de *V. vinifera* est comparable à celle du maïs présentant 1 SNP toutes les 60 bases (Ching et al, 2002) à 104 bases (Tenaillon et al, 2001) et est plus importante que celle de la tomate (1 SNP toutes les 139 bp, Labate et al, 2009), d'*Arabidopsis* (1 SNP toutes les 336 bases; Schmid et al, 2003) ou du blé (1 SNP toutes les 540 bases; Somers et al, 2003). La diversité nucléotidique de la vigne ( $\Pi$ : nombre moyen de différences nucléotidiques par site entre paires de séquences prises au hasard) est de 0.0051 (Lijavetzky et al, 2007) similaire à celle du maïs ( $\Pi$  =0.0063; Ching et al, 2002) et environ 5 fois plus importante que celle du soja ( $\Pi$  =0.0012; Zhu et al, 2003) et de la tomate ( $\Pi$  =0.0010; Labate et al, 2009).

### 1.1.2.1.3. Séquençage du génome de la vigne

Le séquençage du génome du Pinot Noir a permis de mettre en évidence 1 751 176 SNP entre les deux haplotypes ainsi que plus d'un million d'insertion / délétion (INDEL) couvrant 48.8 Mb et 65.1 Mb correspondant à de l'ADN hémizygote (Velasco et al, 2007). En moyenne, 11.2 % de variation nucléotidique sont observés entre la séquence ADN des chromosomes homologues du Pinot noir, en partie lié à la dynamique des éléments transposables. Ces travaux donnent également une image de la répartition des polymorphismes le long du génome. Les régions codantes et non-codantes présentent un degré de polymorphisme différent avec en moyenne 2,5 SNP et 5,5 SNP toutes les kilo bases (kb) respectivement. De plus, les auteurs ont observé des régions présentant des pics de

20 SSR 4370 accessions  
provenant de la plus grande collection mondiale  
de *Vitis vinifera* L.  
(Domaine de Vassal, INRA Montpellier)



**Figure 1.3:** Utilisation des données de diversité génétique et de structure des populations de *V. Vinifera* pour définir des sous-ensembles de génotypes pour différentes applications (d'après Le Cunff et al, 2008 ; Laucou et al, 2011; Nicolas S., Lacombe T., Le Cunff L., Boursiquot JM. en préparation)



polymorphisme, comme pour le chromosome 19 dont certaines régions ont une fréquence moyenne en SNP de 5 et 7,5 par kb alors que le reste du chromosome est moins polymorphe. Le séquençage du Pinot Noir a permis de confirmer la forte hétérozygotie de la vigne, de quantifier et de mettre en évidence la nature des différences entre les deux haplotypes.

Enfin, les nouvelles générations de techniques de séquençage (Nagarajan et Pop, 2010) combinées avec l'existence d'une séquence de référence, permettent maintenant de détecter à très haut débit des variations de type SNP et petits INDEL (< 10bp) dans des collections de génotypes. Ainsi, 12 cultivars de *V. vinifera* (11 *V. vinifera* ssp *sativa* dont le génotype de référence PN40024 et un *V. vinifera* ssp *silvestris*) et six représentants d'autres espèces du genre *Vitis* ont été re-séquencés à l'aide d'une plateforme Illumina. Ceci a permis la découverte de 469 470 SNP (avec une couverture >10 X et couvrant environ 2,3% du génome de référence ; Myles et al, 2010) et la création d'une puce de génotypage de 9K. Cette dernière étude a également confirmé l'état fortement hétérozygote du génome de la vigne et sa forte diversité nucléotidique.

Les études de la diversité génétique chez *V. vinifera* menées à l'aide de marqueurs génétiques ou par re-séquençage ont révélé un niveau d'hétérozygotie global très élevé chez la vigne, tant au sein du compartiment sauvage que cultivé, ainsi qu'une forte diversité génétique, cependant plus importante dans le compartiment cultivé que sauvage. Ces connaissances et ces outils pourraient maintenant être exploités dans des études de génétique d'association ou pour l'identification de traces de sélection le long du génome, comme chez d'autres espèces (Palaisa et al, 2003; Burke et al, 2005; Wright et al, 2005), dans le but d'identifier des loci ou allèles impliqués dans la variation de caractères d'intérêt agronomique.

#### 1.1.2.2. Structure des populations de *Vitis vinifera* L.

La plus grande collection mondiale de *V. vinifera* se trouve sur le domaine de Vassal à Marseillan en France et dépend de l'INRA de Montpellier. Cette collection regroupe 4370 accessions de *V. vinifera*, comprenant des variétés cultivées et des sauvages, et également des représentants des espèces du genre *Vitis*. L'ensemble de ce matériel a été génotypé à l'aide de 20 SSR répartis sur les 19 chromosomes de la vigne, ce qui a conduit à identifier 2323 cultivars uniques chez *V. vinifera* ssp *sativa* (Laucou et al, 2011).

Ces données ont également permis de définir des core-collections emboîtées de 12, 24, 48 et 96 génotypes, maximisant la variabilité génétique du compartiment cultivé (Le Cunff et al, 2008; Figure 1.3). La core-collection 24 représente respectivement 73% et 88% de la diversité

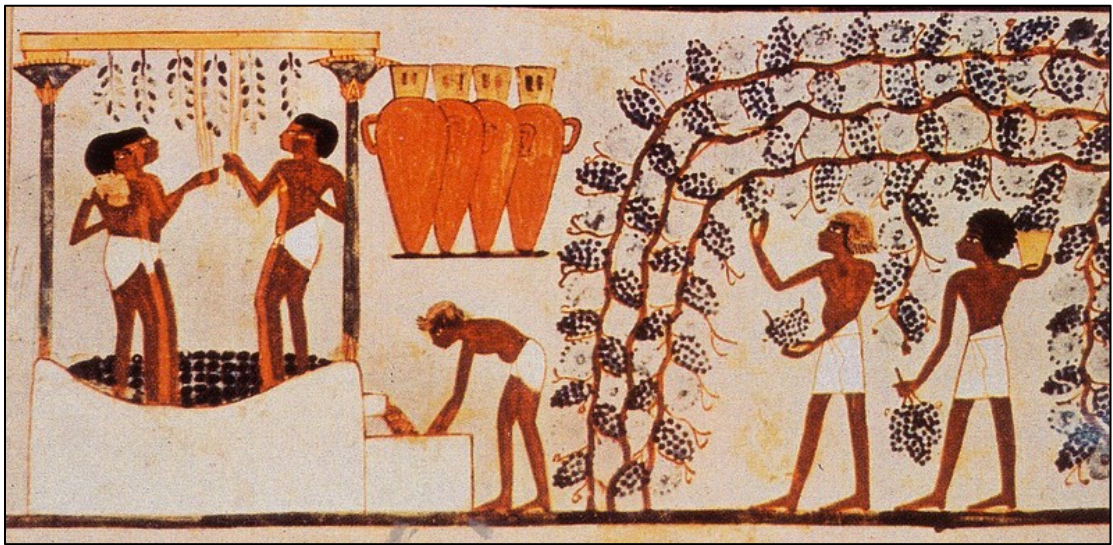
SSR et SNP du compartiment cultivé (Le Cunff et al, 2008; Laucou et al, 2011). Ces core-collections sont un outil précieux pour la recherche de polymorphisme de séquences car elles permettent de détecter un maximum de polymorphismes en utilisant un échantillon de taille raisonnable. C'est dans ce but que la core-collection 24 a été utilisée comme noyau dans la constitution d'un panel de génotypes dans un projet de découverte de SNP dans le genre *Vitis* (projet SNPGrpMap regroupant toutes les unités INRA s'intéressant à la vigne, auquel j'ai participé). A ces 24 individus, ont été ajoutés 3 autres génotypes *V. vinifera* cultivés ayant un intérêt en sélection ou en cartographie génétique, le génotype de référence PN40024 (Jaillon et al 2007), 7 génotypes *V. vinifera* ssp *silvestris* (les moins apparentés disponibles dans la collection du domaine de Vassal) et 12 représentants d'autres espèces du genre *Vitis* (Figure 1.3). Ce projet avait pour objectif de re-séquencer 1100 fragments de gènes de 800 paires de bases (pb) bien répartis sur le génome comprenant un ensemble de gènes candidats pour divers caractères d'intérêt agronomique sur une collection de 48 génotypes, afin d'identifier un maximum de SNP qui pourront être génotypés ultérieurement dans un ensemble plus important de génotypes pour les études de génétique d'association et recherche de trace de sélection. Le même type d'approche est en cours pour le compartiment *V. vinifera* ssp. *silvestris* en se basant sur une grande collecte européenne de génotypes sauvages dans le cadre du projet européen GrapeGen06 (coordonné par R. Bacilieri, UMR AGAP), aboutissant très récemment à la création d'une collection de 96 *V. vinifera* sauvages, comprenant des génotypes peu apparentés, divers géographiquement et maximisant la variabilité du compartiment sauvage telle qu'elle a pu être décrite dans le cadre de ce projet (S. Nicolas, T. Lacombe, J-M. Boursiquot, communication personnelle; Figure 1.3).

Les données de génotypage obtenues sur de larges collections de ressources génétiques permettent bien évidemment avant tout d'étudier la structure des populations chez l'espèce d'intérêt. La vigne ne fait pas exception, la communauté dispose maintenant d'une assez bonne vision de la structure des populations de *V. vinifera* ssp *sativa* (Aradhya et al, 2003; Le Cunff et al, 2008; Myles et al, 2011; Laucou et al, 2011). En effet, la vigne cultivée se structure en 3 populations, reflétant son origine géographique et son utilisation : l'ensemble des variétés de Cuve originaire de l'Europe de l'Est (CE), des variétés de Cuve originaire de l'Europe de l'Ouest (CO) et des variétés de Table originaires de l'Europe de l'Est (TE) (Laucou et al, 2011). La structuration du compartiment sauvage est quant à elle en cours de caractérisation dans le cadre du projet GrapeGen06 mais un grand effort reste à faire pour les autres espèces du genre *Vitis*.

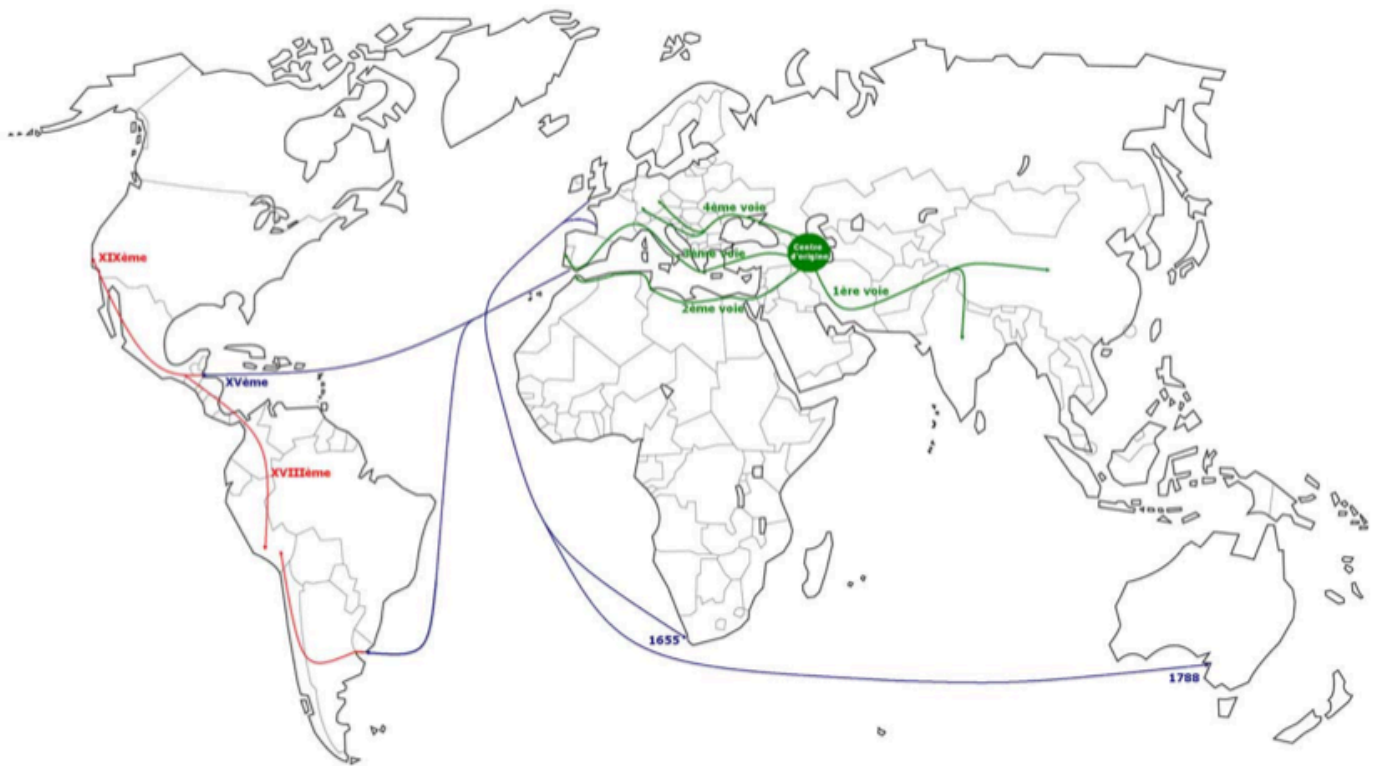
La connaissance de la structure des populations est indispensable dans les études de génétique d'association afin de limiter les faux positifs. En effet, au sein d'échantillons très diversifiés, on observe régulièrement que les individus les plus proches phénotypiquement sont également les plus proches génétiquement, et pas seulement pour des loci responsables de variations phénotypiques. Cette observation conduit fréquemment à considérer des sites neutres comme associés à une variation. Ainsi, tous les individus issus d'une population particulière sont plus susceptibles de partager des éléments de variations phénotypiques et génétiques: c'est ce qu'on appelle la structure de l'échantillon. Il est possible de considérer les individus comme issus d'un mélange de ces populations ancestrales: chaque individu se voit donc assigné un degré d'appartenance à chacune des populations ancestrales (Pritchard et al, 2000). Toutefois, l'appartenance à des groupes ancestraux ne prend en compte que de façon lointaine les relations de proximité génétique entre deux individus pris au hasard dans une population. Une autre contribution significative au contrôle des risques de faux positifs, est d'utiliser directement les relations entre individus par le biais de calcul d'identité par descendance, ce qui conduit à définir une matrice d'apparentement deux à deux (*Kinship*:  $K$ ; Yu et Buckett et al, 2006). Le calcul de la structure de la population et de la matrice d'apparentement repose sur les données de génotypage à l'aide de marqueurs neutres qui sont désormais abondants chez la vigne. Les modèles de génétique d'association prenant comme variables ces deux paramètres apparaissent comme les plus puissants (Zhao et al, 2007; Myles et al, 2009).

L'ensemble des connaissances sur la structure des populations et le degré d'apparentement entre individus a été valorisé pour construire un échantillon destiné à des études de génétique d'association chez la vigne (S. Nicolas, T. Lacombe, L. Le Cunff, communication personnelle; Figure 1.3). Il se compose de 279 génotypes cultivés, structurés en 3 sous-populations: 93 variétés CO, 93 variétés CE, 93 variétés TE (Figure 1.3). Chaque génotype a un pourcentage d'appartenance  $> 80\%$  à l'une des populations, et les génotypes à l'intérieur de chaque sous-population ont été choisis de manière à être les moins apparentés et en privilégiant les génotypes ancestraux. La variabilité génétique de cette collection a été estimée à l'aide de 20 marqueurs SSR et représente plus de 80% de la variabilité du compartiment cultivé, et c'est cet échantillon que j'ai utilisé pour mes études de génétique d'association.

Il est également nécessaire d'avoir une bonne vision de l'histoire évolutive des populations cultivées et sauvages, ainsi que de la structure du Déséquilibre de Liaison (DL) le long du génome pour conduire au mieux des études de génétique chez la vigne.



**Figure 1.4:** Fresque décrivant les vendanges, Thèbes, tombe de Nakht (1590-1390 avant JC).



**Figure 1.5 :** Diffusion de la vigne cultivée d'après Fregoni (1991). Les quatre voies principales sont illustrées en vert (1<sup>ère</sup> voie au Moyen Orient en 3000 av JC, 2<sup>ème</sup> voie par la Palestine et l'Égypte en 4000 av JC, 3<sup>ème</sup> voie par la Méditerranée septentrionale 600 avant JC et 4<sup>ème</sup> voie suivant le courant danubien). Les voies de diffusion plus récentes (maritimes en bleu ou terrestres en rouge) ont finalement contribué à une extension de la vigne dans le Nouveau Monde.

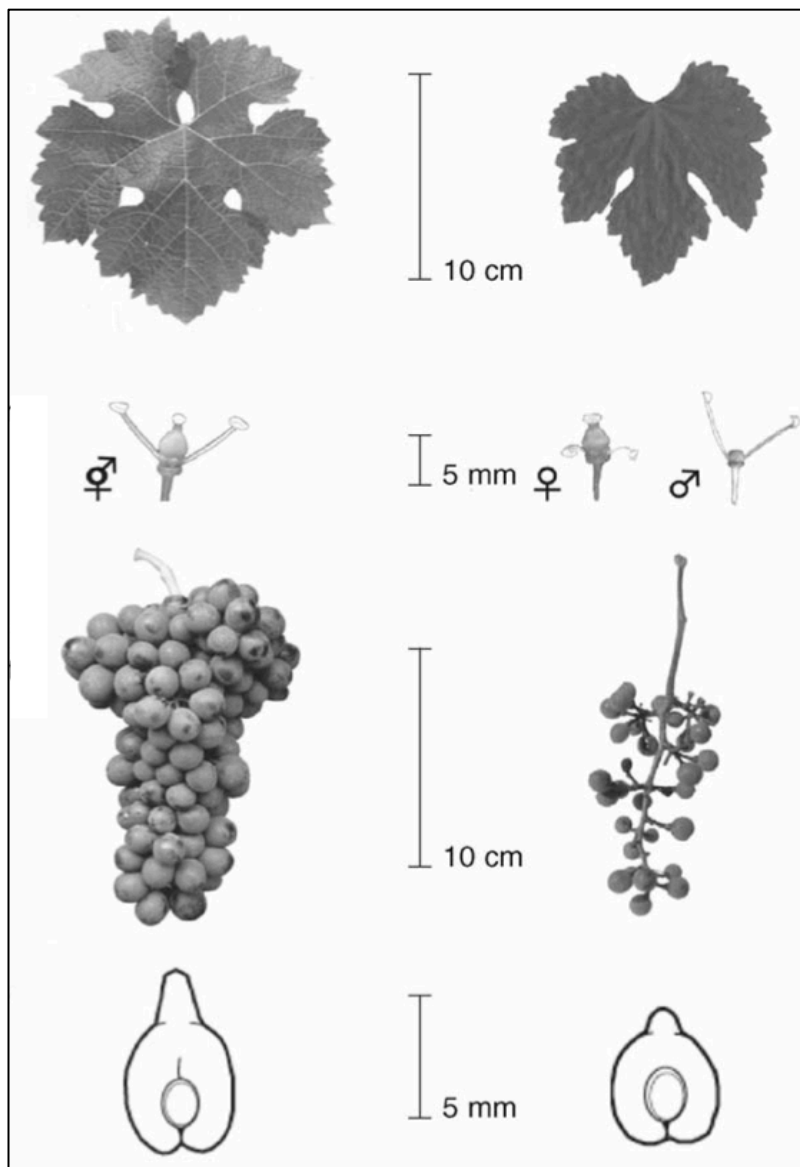
### **1.1.3. Domestication et sélection de la vigne**

#### **1.1.3.1. De la domestication à la viticulture moderne: petit historique**

Depuis la dernière glaciation, la vigne a une aire de répartition qui s'étend du Caucase à l'Himalaya, où elle a vraisemblablement été cueillie et consommée par les populations humaines depuis longtemps. Selon les auteurs, *V. Vinifera* aurait été domestiquée entre 9000 et 4000 ans avant Jésus Christ (JC) (Levadoux, 1956; Marinval, 1997). Les premières traces de consommation consistent en des résidus de vin retrouvés dans une poterie en Iran qui daterait de 5400 à 5000 avant JC (Mc Govern, 2003). Cependant, les premiers témoignages incontestables d'une viticulture organisée nous viennent de l'Égypte ancienne vers 3000 avant JC (Figure 1.4). L'histoire de la vigne et de la viticulture est marquée par une succession d'expansions liées à la diffusion du vin (Figure 1.5). Les Phéniciens et les Grecs ont introduit la viticulture dans les civilisations de l'Ouest méditerranéen pour la partie occidentale (Mc Govern, 2003), puis sous l'influence de l'Empire romain, la vigne s'est étendue plus largement à travers l'Europe vers la Hongrie et la Franconie (Fregoni, 1991). Avec le christianisme, le vin intègre le sacré et accompagne la christianisation jusqu'aux limites les plus septentrionales (Sud de l'Angleterre) et orientales (Oural) (Royer, 1988). De l'autre côté de la Méditerranée, le raisin est plutôt consommé sous forme de fruits frais ou secs et s'installe durablement du Maroc jusqu'aux confins de la Chine (Royer, 1988). Plus récemment, des vignobles furent établis dans les nouveaux mondes (Amériques, Afrique du Sud et Australie) avec les colons des XVII<sup>ème</sup> et XVIII<sup>ème</sup> siècles (Royer, 1988). De nos jours, *V. vinifera* est le deuxième fruit le plus cultivé au monde, sa culture est présente sur tous les continents hormis l'Antarctique et s'étend sur environ 8 millions d'hectares à travers le monde dont 852 000 hectares en France d'après l'Organisation Internationale de la Vigne et du Vin (OIV 2008).

#### **1.1.3.2. La domestication**

La domestication semble avoir concerné principalement deux caractères distincts (Figure 1.6; This et al, 2006). En premier lieu, une modification des organes reproducteurs de formes dioïques (chez les sauvages) à hermaphrodites (chez les cultivés) a permis de s'affranchir des aléas de l'allopollinisation et de ne cultiver que des vignes productives. D'autre part, la domestication a modifié les caractères morphologiques des baies permettant



**Figure 1.6:** Principales différences morphologiques entre la vigne sauvage (*V. vinifera* L. ssp. *silvestris*, à droite) et cultivée (*V. vinifera* L. ssp. *sativa*; à gauche) au niveau des feuilles, fleurs, grappes à maturité et pépins (This et al, 2006).

une production régulière, un meilleur rendement (grappes plus compactes et fruits plus gros: Figure 1.6) et une amélioration de la qualité organoleptique (augmentation des sucres pour une meilleure fermentation; Pouget, 1988) mais a également eu des conséquences indirectes sur la forme des feuilles et des pépins (Figure 1.6; This et al, 2006). Ce dernier caractère est utilisé dans les analyses de vestiges archéologiques pour différencier les vignes sauvages des cultivées (Terral et al, 2010). La comparaison morphologique des vignes sauvages et de l'ensemble des variétés cultivées suggère l'existence d'un gradient de domestication au sein des cépages. Par exemple, d'après Levadoux (1956), certains cépages de cuve comme le Pinot et le Petit Verdot (également appelé Lambrusquet) présentent des caractères morphologiques proches des lambrusques: petit gynécée, petits pépins ronds, petites grappes à petites baies rondes. Des formes plus évoluées, comme chez certaines variétés de raisins de table, telles que le Dattier de Beyrouth ou le Muscat d'Alexandrie, se distinguent par des grappes plus grosses, des pépins piriformes et de grosses baies de forme, de saveur et de couleur variables. De plus, plusieurs études utilisant des marqueurs microsatellites ou SNP ont montré une distinction claire au niveau génétique entre le compartiment sauvage et cultivé (This et al, 2001; Lacombe et al, 2002; Grassi et al, 2003; Aradhya et al, 2003; Snoussi et al, 2004; Arroyo-Garcia et al, 2006; Vezzulli et al, 2008; Laucou et al, 2011).

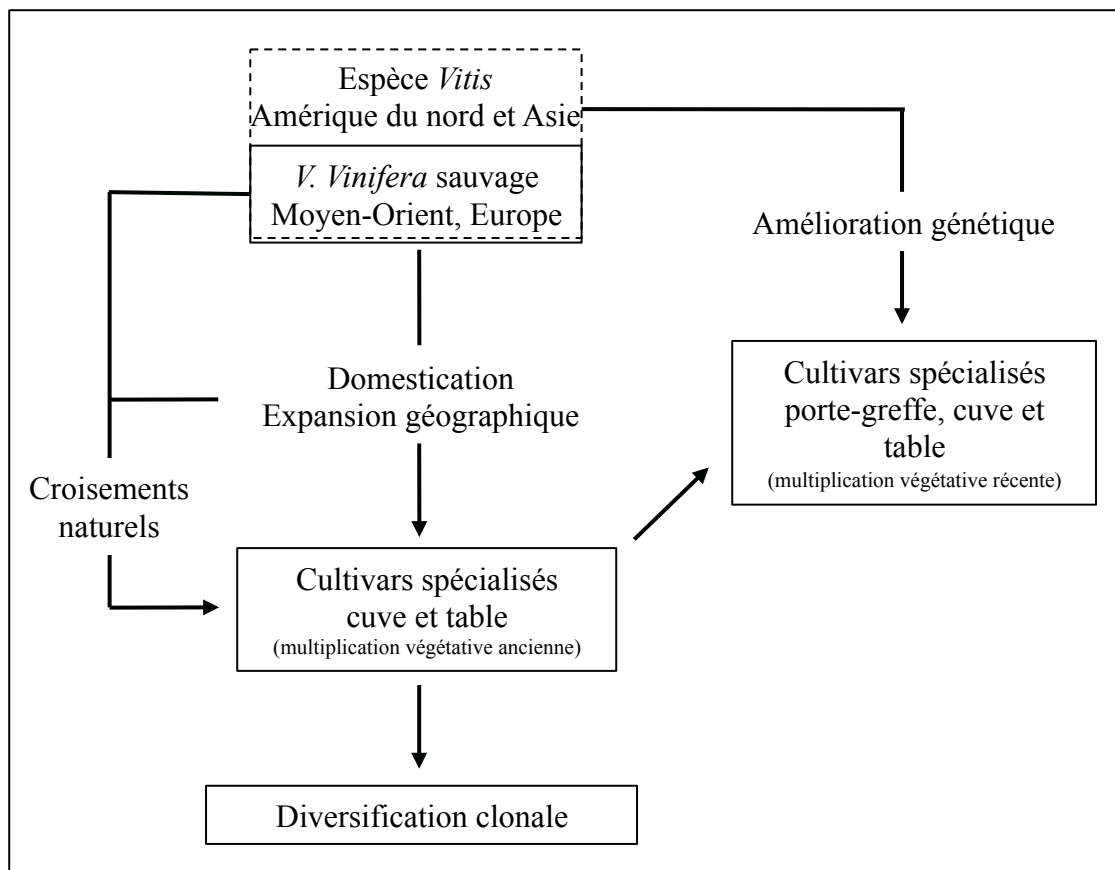
Le nombre d'événements de domestication et leur localisation géographique sont encore discutés. Deux hypothèses divergentes ont été émises. La première consiste en la domestication d'un nombre limité de vignes sauvages en une seule localisation suivie d'une diffusion des cultivars à d'autres régions (Olmo, 1976). La deuxième hypothèse repose sur de multiples événements de domestication ayant impliqué un grand nombre de vignes sauvages pendant une longue période sur l'aire entière de distribution des vignes sauvages (Mullins et al, 1992). En accord avec la première hypothèse, une étude récente portant sur l'analyse de la diversité haplotypique et de la structure d'une population de 950 vignes cultivées et 59 vignes sauvages, de diverses provenances géographiques, (Myles et al, 2011) a révélé à partir des données de génotypage de 5387 SNP répartis dans tout le génome, que l'ensemble des vignes cultivées était plus proche génétiquement des populations sauvages de l'Est que de celles de l'Ouest. Cette étude conforte l'hypothèse d'un centre unique de domestication localisé au Moyen-Orient. Cependant, de nombreux génotypes sauvages de cette étude pourraient être des hybrides entre des *V. vinifera* sauvages et cultivées ou bien leur appartenance au compartiment sauvage n'est pas certaine (Myles et al, 2011). Ces conclusions sont donc à prendre avec précaution. D'un autre côté, l'existence de différences morphologiques entre les



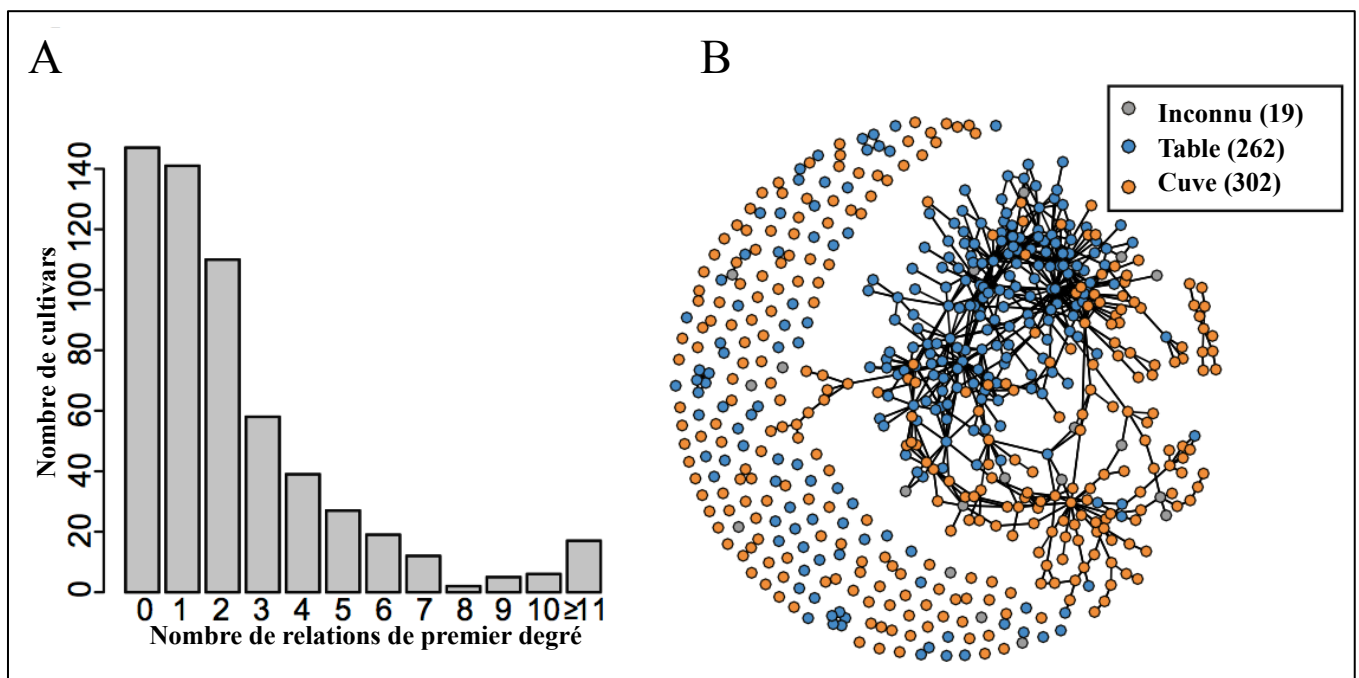


cultivars de l'Est et de l'Ouest de l'Europe (à l'Est principalement des cultivars de table avec de grosses baies et à l'Ouest des cultivars de cuve avec de plus petites baies; Negrul, 1938; Levadoux, 1956) est à l'origine de la seconde hypothèse, selon laquelle le compartiment sauvage local aurait contribué significativement à la création des variétés, certainement à travers de multiples événements de domestication et de sélection des vignes sauvages locales (Mullins et al, 1992). Une étude portant sur l'analyse de la diversité haplotypique de l'ADN mitochondrial de 1201 *V. vinifera* (688 sauvages et 513 cultivées) réparties sur l'aire géographique de *V. vinifera* (l'Europe, l'Afrique du nord, le Proche Orient et le Moyen-Orient) met en évidence l'existence d'une contribution génétique des populations sauvages de l'Est comme de l'Ouest dans la structuration des populations de vignes cultivées, suggérant donc plutôt deux événements indépendants de domestication: un en Europe de l'Est et un en Europe de l'Ouest (Arroyo-Garcia et al, 2006). Deux hypothèses peuvent être émises pour expliquer la proximité génétique plus forte entre les variétés cultivées de l'Ouest et les vignes sauvages de l'Ouest: soit les vignes sauvages d'Europe de l'Ouest ont subi des introgressions génétiques provenant de vignes cultivées de la même région géographique, soit il y a bien eu une domestication locale des vignes sauvages de l'Ouest à l'origine des vignes cultivées de l'Ouest. La seconde hypothèse a été confirmée par plusieurs études, montrant que les populations sauvages originaires d'Europe de l'Ouest ne présentent pas d'introgression de gènes provenant de vignes cultivées de la même région (This et al, 2001; Snoussi et al, 2004; Di Vecchi-Staraz et al, 2009) suggérant donc une domestication locale des vignes sauvages de l'Ouest. La difficulté de ce type d'analyse est de disposer d'un échantillon représentatif des deux compartiments, sauvage et cultivé. Dans le compartiment cultivé, la notion de provenance géographique peut s'avérer très ambiguë en raison de nombreux effets de migration alors que le compartiment sauvage a été très affecté par la crise phylloxérique conduisant à une réduction importante de diversité génétique et reste encore mal connu.

Une autre manière d'étudier la domestication consiste à identifier les gènes ciblés pendant ce processus par des approches de détection de QTL dans des populations issues de croisements entre sauvages et cultivées (Purugganan et Fuller, 2009; Gross et Olsen, 2010 pour revues). De nombreux gènes impliqués dans la transition des espèces sauvages aux variétés cultivées, tels que les gènes *tb1* et *tga1* responsables respectivement de l'absence de ramifications secondaires (Doebley et al, 1997) et du durcissement des grains chez le maïs (Wang et al, 2005), les gènes *sh4*, *qSH1* chez le riz responsables de l'égrainage (Li et al, 2006; Konishi et al, 2006), les gènes *Rc* et *PROG1* responsables de la couleur et du nombre



**Figure 1.7:** Schéma simplifié des étapes de domestication et de sélection de la vigne d'après Olmo (1976).



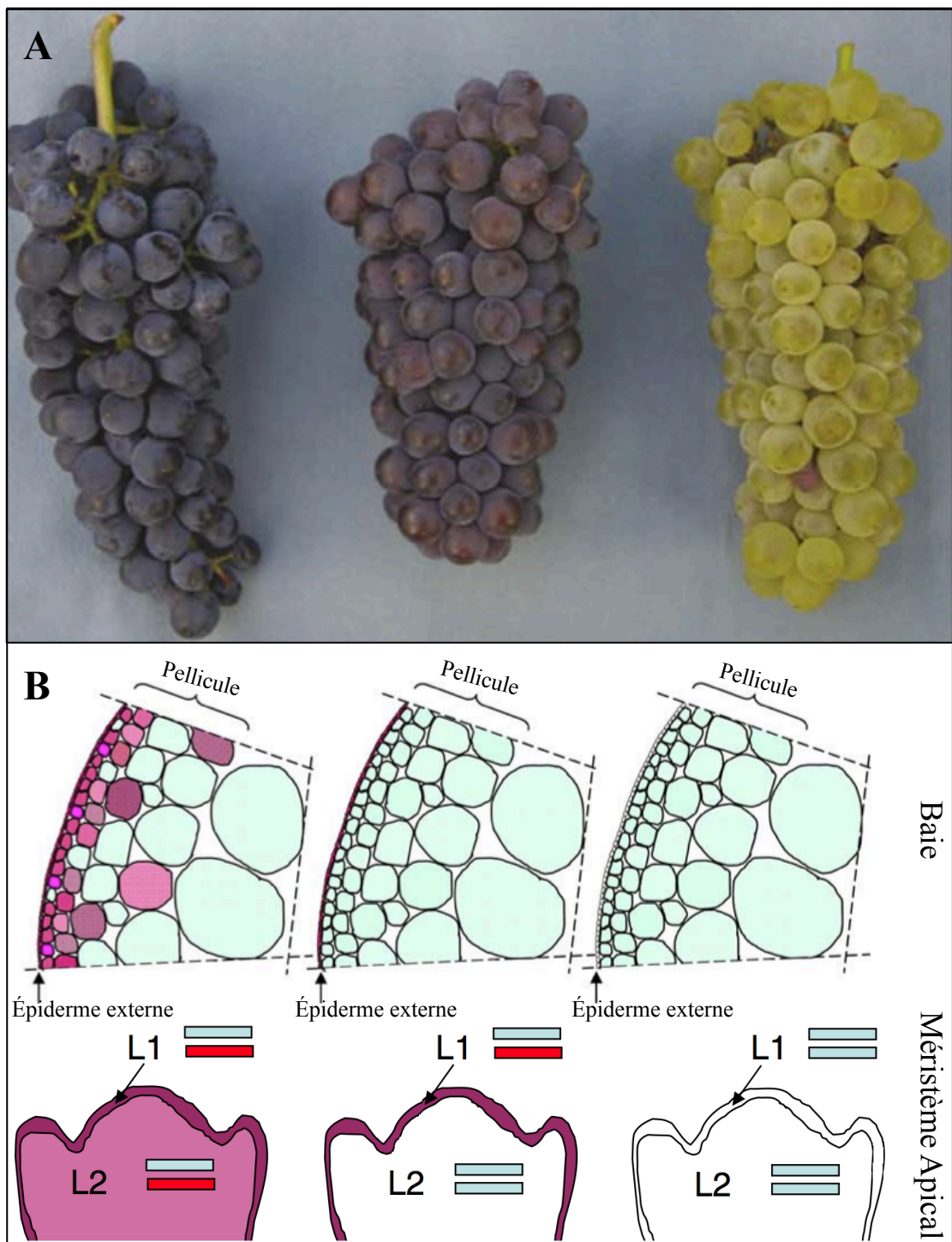
**Figure 1.8:** Relations de premier degré d'apparement entre 583 génotypes *V. vinifera* cultivés uniques provenant de la collection de l'USDA (USA; Myles et al 2011). **(A)** Distribution du nombre de relations de premier degré par génotype. **(B)** Structure généalogique chez *V. vinifera* ssp *sativa* représentée sous la forme d'un réseau. Un seul réseau interconnecté est clairement visible et comprend 58% des cultivars qui sont reliés entre eux par une série de relations de premier degré.

de grains chez le riz (Sweeney et al, 2007; Tan et al, 2008) et le gène *fw2.2* impliqué dans la variation de la taille du fruit chez la tomate (Frary et al, 2000) ont ainsi été identifiés durant les 20 dernières années. Ces découvertes ont permis de mieux comprendre le mécanisme de la domestication chez les espèces en question. Chez la vigne, le gène impliqué dans le déterminisme du sexe a été récemment isolé et est en cours de validation fonctionnelle (L. Torregrosa communication personnelle). Il s'agira du premier gène de ce type identifié chez cette espèce.

### 1.1.3.3. La sélection

Des processus de sélection ont suivi la domestication et sont à l'origine de l'importante diversité morphologique de la vigne cultivée actuelle (Olmo, 1976; Figure 1.7). Cette sélection s'appuie sur deux principaux mécanismes générateurs de variabilité.

Tout d'abord, la reproduction sexuée semble avoir joué un rôle important dans les premières étapes suivant la domestication, durant les phases d'expansion de la viticulture (Bowers, 1999), conduisant à la création de nouvelles variétés adaptées à chaque environnement et aux utilisations finales: raisins de cuve, raisins de table ou raisins secs. Les cépages de table se caractérisent par des grosses baies pulpeuses parfois apyrènes, tandis que les cépages de cuve ont des baies plus petites, juteuses et contenant des pépins. Par exemple, le Cabernet-Sauvignon, le Chardonnay ou la Syrah sont les descendants directs de croisement entre des variétés très anciennes (Bowers et al, 1999 et 2000; Boursiquot et al, 2009). Plus généralement, l'étude du profil génotypique de plusieurs marqueurs SSR ou SNP dans de larges collections de variétés de vigne montre que l'ensemble des variétés actuelles est structuré en groupes de génotypes apparentés, souvent au premier degré (Bowers et Meredith, 1997; Bowers et al, 1999; Vouillamoz et Grando, 2006; Cipriani et al, 2010; Myles et al, 2011; Figure 1.8). Cela confirme le faible nombre de générations à l'origine des génotypes cultivés de nos jours. De fait, la sélection d'un type particulier est un long processus, si on considère que la période juvénile pendant laquelle la vigne ne produit pas encore de fruits qui est de 3 à 5 ans et le temps additionnel pour évaluer les caractères importants pour la production de vin (This et al, 2006). La découverte de la multiplication végétative par marcottage (enracinement des rameaux sans que ceux-ci ne soient séparés de la plante-mère) a facilité la pérennisation et la diffusion des génotypes ayant les structures génétiques les plus performantes dès l'époque romaine. Certains cultivars comme le Pinot Noir et le Gouais auraient ainsi été maintenus par propagation végétative, des Romains à nos jours (Bowers et al 1999).



**Figure 1.9:** Variation clonale dans la famille des Pinot. Photo (A) de clones de Pinot se distinguant par la couleur de la pellicule: Pinot noir (à gauche), Pinot gris (au milieu) et Pinot blanc (à droite; d'après This et al, 2006) et (B) structure chimérique ou non de la baie : les barres rouges et blanches correspondent à l'état allélique du gène *VvMYBA1* codant pour un facteur de transcription régulant la voie de biosynthèse des anthocyanes dans les différentes couches cellulaires de la baie (Walker et al, 2006; Yakushiji et al, 2006) .

En second lieu, en l'absence de reproduction sexuée, les mutations somatiques représentent l'unique source de variation génétique chez une plante à multiplication végétative. Les mutations somatiques sont relativement fréquentes chez la vigne, faciles à propager et ont fortement contribué à la diversité du compartiment cultivé (Olmo, 1935; Snyder et Harmon, 1935). De nombreux variants clonaux se distinguant par leur taille, leur couleur, l'acidité de la baie ou la présence de pépins ont ainsi été identifiés, comme par exemple chez la Sultanine (Olmo, 1934; Boubals et al, 1971; Coombe, 1960; Boss et al, 1996). L'exemple le plus célèbre est la famille des Pinot: Pinot blanc, Pinot noir, Pinot gris (variation de la couleur de la baie; Figure 1.8) mais aussi Pinot meunier (de la morphologie de la feuille), Pinot fin (du rendement), Pinot droit (du port de la plante), Pinot Moure (de l'absence de cire sur la baie)... L'utilisation des marqueurs microsatellites a permis de mettre en évidence l'importance de ce phénomène, qui ne conduit pas toujours à un phénotype visible (Franks et al, 2002; Hocquigny, 2003; Hocquigny et al, 2004; Pelsy, 2010; Pelsy et al, 2010).

Chez la vigne, le développement végétatif s'effectue à partir d'un méristème apical stratifié, constitué de plusieurs couches cellulaires proliférant de manière plus ou moins indépendante (Klekowski et al, 1985). Les couches L1 vont produire l'épiderme alors que les couches L2 sont à l'origine du parenchyme, des tissus gamétophytiques et vasculaires contribuant aux différents tissus et organes de la plante (Figure 1.9; Thompson and Olmo, 1963). Lorsqu'une mutation affecte une cellule du méristème apical, cette cellule mutante se divise par mitose et produit un secteur muté qui se différenciera génétiquement des autres secteurs de la plante. De tels états sont appelés "chimériques" et sont fréquents chez la vigne. Les mutations ainsi créées peuvent s'accumuler au fil du temps menant à des différences morphologiques et à la définition de nouvelles variétés (This et al, 2006). La reproduction végétative permet le maintien et la propagation de ces mutations dans leur état chimérique, ce qui n'est pas toujours possible *via* la reproduction sexuée. En effet, les mutations somatiques ne peuvent être propagées par reproduction sexuée que si la mutation touche la couche cellulaire à l'origine des gamètes (Pelsy, 2010).

Chez une plante pérenne comme la vigne où les mutants artificiels sont difficiles à obtenir (Botta et al, 1987; Kuksova et al, 1997; Khawale et al, 2007), les mutants spontanés représentent actuellement une ressource unique pour comprendre le contrôle génétique de systèmes développementaux complexes comme la formation du fruit. Même si la mutation causale est souvent présente sous une forme chimérique compliquant les analyses génétiques, plusieurs études ont permis d'identifier les mutations somatiques de type SNP ou élément

**Tableau 1.3:** Liste des mutations somatiques permettant la distinction morphologique entre clones d’une même variété dont le gène muté a été identifié.

Variété	Phénotype	Couche cellulaire mutée	Gène	Fonction putative	Type de mutation	Caractère de la mutation	Reference
Pinot Noir	baie de couleur blanche	L1	<i>VvMYBA1</i>	enzyme de la voix de biosynthèse des anthocyanes	élément transposable	récessive	Kobayashi et al, 2006 Walker et al, 2006
	baie de couleur rose	L1 et L2					
Carignan	ramification extrême de la grappe	L2	<i>VvTFL1A</i>	determinisme de la structure de l'inflorescence	élément transposable	dominante	Fernandez et al, 2010
Pinot Meunier	augmentation de la pilosité des feuilles	L1	<i>VvGAI</i>	régulateur négatif de la signalisation à l'acide gibbérélique	SNP	dominante	Franck et al, 2002
	nanisme et developpement d'inflorescences à la place de vrilles	L1 et L2					Boss et Thomas, 2002

transposable (TE), à l'origine de différences morphologiques entre clones d'une même variété (Boss et Thomas, 2002; Kobayashi et al, 2006; Walker et al, 2006 ; Fernandez et al, 2010) et sont répertoriées dans le tableau 1.3.





## 1.2. Analyse du déterminisme génétique et moléculaire de la variabilité d'un caractère

Les études génétiques chez la vigne sont maintenant facilitées par l'existence de nombreuses ressources publiques dont beaucoup ont été développées dans le cadre du consortium international du génome de la vigne (*International Grape Genome Program* IGGP; [www.vitaceae.org](http://www.vitaceae.org)). En particulier, de nombreuses cartes génétiques ont été construites chez la vigne (Adam-Blondon et al, 2011) à l'aide de marqueurs SSR ou SNP, les plus complètes étant celles qui ont servi à ordonner les super-contigs de séquences dans le cadre des projets de séquençage du génome:

- trois cartes génétiques contenant des marqueurs transférables ont été construites, une carte issue du croisement Syrah x Grenache comptant 325 marqueurs SSR, une autre issue du croisement Chardonnay x Bianca comptant 379 marqueurs SSR et une carte consensus comprenant l'ensemble des marqueurs, soit 514 marqueurs SSR (Doligez et al, 2006);
- une carte génétique issue du croisement Syrah x Pinot Noir qui compte 1767 marqueurs moléculaires couvrant 1276 cM répartis sur les 19 groupes de liaison (Troggio et al, 2007; Velasco et al, 2007).

Deux séquences du génome de *V. vinifera*, obtenues par la stratégie Whole Genome Shotgun, ont été publiées en 2007:

- une séquence de référence de haute qualité obtenue à partir de la lignée PN40024 quasi-homozygote dérivée du Pinot Noir (< 7 % d'hétérozygotie résiduelle; Jaillon et al, 2007). Au début de ma thèse, la version couvrant 8,4X le génome (1X de séquence est égal à 1 équivalent génome, soit 487 Méga bases (Mb) chez la vigne) était disponible avec 69% des super-contigs de séquence ancrés sur la carte génétique. En 2009, une nouvelle version couvrant 12X le génome et avec 91 % des super-contigs ancrés sur la carte génétique a été mise en accès libre;
- une séquence du génome hautement hétérozygote du Pinot Noir clone 115 (Velasco et al, 2007) couvrant 10,7X le génome et avec 81% des super-contigs de séquence ancrés sur la carte génétique.

Les annotations automatiques sont également disponibles pour chacune des versions des séquences de génome. Ces annotations ont été générées par le logiciel Gaze (Howe et al, 2002) intégrant plusieurs couches de données (prédictions de gènes *ab initio*, alignement avec des marqueurs de séquences exprimées (*Expressed Sequence Tag* ou EST), des ADN



complémentaires (ADNc), ou des protéines connues du règne végétal. Une deuxième annotation automatique a été réalisée à l'aide du logiciel Eugene par l'URGV (Samson et al, 2004) pour les deux versions de la séquence de référence, et était disponible dans la base de données Flagdb<sup>++</sup> (<http://urgv.evry.inra.fr/projects/FLAGdb++/HTML/index.shtml>). Le nombre de modèles de gènes prédits par ces différents logiciels sont différents 26 346 pour Gaze et 44 417 pour EuGene, reflétant le caractère non définitif de ces annotations qui doivent ensuite être affinées par les communautés scientifiques au cours du temps (Martin et al, 2010).

La mise à disposition publique de la séquence du génome de la vigne et de son annotation automatique est un outil précieux permettant de grandes avancées notamment dans l'identification de gènes impliqués dans des caractères d'intérêt agronomique. En effet, de nombreux caractères importants en sélection, comme le rendement, la précocité de floraison ainsi que certaines résistances aux pathogènes, résultent de l'effet de la ségrégation de plusieurs gènes et de leur interaction avec l'environnement (East, 1910), et le poids de la baie chez la vigne fait partie de ces caractères.

Plusieurs stratégies peuvent être utilisées afin d'identifier les loci impliqués dans la variation d'un caractère quantitatif d'intérêt agronomique. Certaines permettent de cribler le génome entier sans *a priori*, d'autres une région particulière ou des gènes candidats pour le caractère d'intérêt.

### **1.2.1. Détection de QTL dans des populations issues de croisements contrôlés**

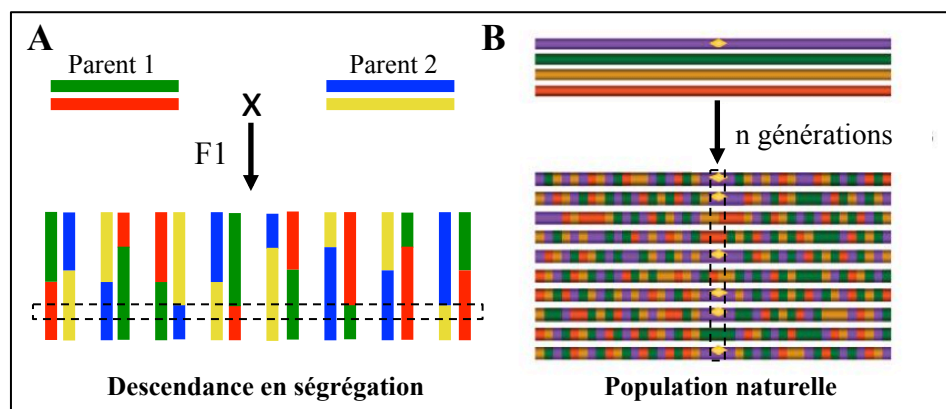
La détection de régions impliquées dans des caractères quantitatifs (*Quantitatif Trait Loci*: QTL) est une méthode qui cribble le génome entier sans *a priori*. Elle repose sur une approche de cartographie génétique qui s'appuie sur la notion de liaison génétique. En l'absence de recombinaison, l'intégralité des marqueurs génétiques présents sur un chromosome, donc physiquement liés, est transmis d'un seul tenant à la descendance. Cependant, lors de la méiose, des recombinaisons peuvent survenir *via* les crossing-over, et conduire à des événements de réarrangements entre deux marqueurs sur un même chromosome qui peuvent être comptés (Morgan, 1911). La distance qui sépare deux loci le long d'un chromosome est directement proportionnelle au nombre de ces recombinaisons. En comptant la fréquence de ces recombinaisons, on peut donc calculer la distance relative

**Tableau 1.4:** Détections de QTL réalisées chez la vigne pour des caractères d'intérêts agronomiques.

Caractère	Type de descendance	Nombre d'individu	Groupe de liaison <sup>§</sup>	Référence
Sexe de la fleur	F1*	58	2	Dalbó et al, 2000
	F1*	188	2	Lowe et al, 2006
	F1*	181	2	Riaz et al, 2006
	F1*	138	2, (6, 7, 10)	Marguerit et al, 2009
Couleur de la baie	F1	139	2	Doligez et al, 2002
	F1	153	2	Fischer et al, 2004
	F1	144	2	Mejia et al, 2007
	F1*	89	2	Salmaso et al, 2008
	F1	163	2	Constantini et al, 2008
Poids/ taille de la baie	F1	139	18	Doligez et al, 2002
	F1	184	(4, 5, 13, 16, 20)	Fanizza et al, 2005
	F1	118	18, (15)	Cabezas et al, 2006
	F1	144	18, (15)	Mejia et al, 2007
	F1	163	18, (1, 12)	Constantini et al, 2008
	F1	139	18, (2, 8)	Mejia et al, 2011 et résultats non publiés
Phénologie de la baie	F1	163	5, 10, (1, 2, 7, 10)	Battilana et al. 2009
	F1	174	5, (2, 12, 10, 15)	Battilana et al. 2009
	autofécondation	121	5, (1, 10, 13, 15)	Duchêne et al. 2009
	F1*	153	7, 8, 16	Fischer et al. 2004
	F1	144	18	Mejia et al. 2007
Apyrénie	F1	139	18	Lahogue et al, 1998
	F1	139	18	Doligez et al, 2002
	F1	118	18, (1, 3, 10, 14)	Cabezas et al, 2006
	F1	144	18, (4, 16)	Mejia et al, 2007
	F1	163	18	Constantini et al, 2008
	F1	139	18	Mejia et al, 2011
	F1	163	1, 2, 6, 16	Costantini et al. 2008
Résistance aux maladies	F1*	157	12	Pauquet et al, 2001; Donald et al, 2002
	F1*	310	13	Hoffman et al. 2008
	F1*	151	13	Coleman et al.2009
	F1*	153 / 144	15	Fischer et al. 2004 ; Welter et al, 2007
	F1*	139	12	Merdinoglu et al, 2003
	F1*	153 / 144	18, (5, 4)	Fischer et al. 2004 ; Welter et al, 2007
	F1*	138	9, 12	Marguerit et al, 2009
	F1*	73	14	Kim et al, 2008
	F1*	188	19	Xu et al, 2008
	F1*	188	13	Zhang et al, 2009

<sup>§</sup> les QTL mineurs sont entre parenthèses

\* croisement entre *V. vinifera* et espèce *Vitis* ou hybride interspécifique



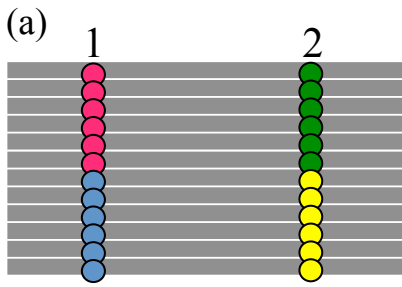
**Figure 1.10:** Comparaison des populations utilisées (A) en cartographie de QTL et (B) dans des études de génétique d'association. Les différentes couleurs symbolisent les différents allèles et la taille des segments est proportionnelle au nombre de recombinaisons accumulées (d'après Zhu et al, 2008).

(appelée centiMorgan: cM) de ces deux marqueurs liés génétiquement. La cartographie QTL consiste à rechercher des corrélations entre un phénotype et les polymorphismes de marqueurs génétiques bien répartis sur le génome, sur une descendance en ségrégation pour le caractère d'intérêt. Plus les marqueurs sont proches du QTL, plus ils seront liés au phénotype du caractère d'intérêt dans la descendance. Différents types de descendance peuvent être utilisés, les plus courants chez les plantes sont des populations F<sub>2</sub>, des populations de lignées recombinantes (*Recombinant Inbred Line*: RIL) ou des lignées d'haploïdes doublés. Le choix de la descendance dépend du modèle biologique étudié. Dans le cas de la vigne, plante pérenne très hétérozygote qui est sensible à l'homozygotie, la plupart des populations étudiées correspondent à des populations de pleins frères directement issues du croisement entre deux génotypes en général hautement hétérozygotes (Tableau 1.4 et Figure 1.10; Weeden, 1994; Grattapaglia et al, 1996), et plus rarement, à des populations dérivées de l'autofécondation d'un individu (Duchêne et al, 2009). Dans le premier cas, deux cartes parentales sont construites sur la base des marqueurs hétérozygotes chez chacun des parents et la détection de QTL est réalisée indépendamment chez chacun des parents. Une carte dite consensus peut également être établie en utilisant les marqueurs communs aux deux cartes parentales.

Chez la vigne, de nombreux QTL sont maintenant cartographiés (Tableau 1.4), comme par exemple le locus responsable du déterminisme du sexe (Dalbo et al, 2000; Lowe and Walker, 2006; Riaz et al, 2006; Marguerite et al, 2009), des QTL pour le poids des baies et l'apyrénie (Doligez et al, 2002; Fischer et al, 2004; Mejia et al, 2007; Salmaso et al, 2008; Costantini et al, 2008), des QTL de résistance aux maladies (Pauquet et al, 2001; Donald et al, 2002; Merdinoglu et al, 2003; Krivanek et al, 2006; Welter et al, 2007; Hoffman et al, 2008; Xu et al, 2008; Coleman et al, 2009; Marguerit et al, 2009; Zhang et al, 2009) et des QTL pour des caractères en relation avec la phénologie ou le développement (Fischer et al, 2004; Mejia et al, 2007; Costantini et al, 2008 ; Battilana et al, 2009; Duchêne et al, 2009).

Plus généralement, bien que la cartographie de QTL continue à être une stratégie de choix pour identifier les gènes responsables des variations quantitatives chez les plantes, car elle permet théoriquement d'ajuster le plan expérimental (nombre d'individus en ségrégation, nombre de marqueurs, nombre de répétitions pour le phénotypage) aux besoins de la précision de détection, ces ajustements peuvent s'avérer très lourds chez les plantes pérennes en termes de temps (pour l'établissement de populations en ségrégation, phénotypage possible une fois par an pour les caractères de la baie et période juvénile de 3 à 5 ans) et d'espace (taille des individus). Pour ces raisons, la taille des populations étudiées n'excède que très rarement 200

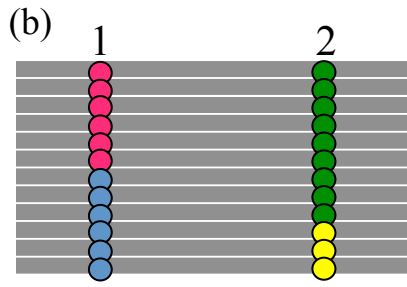
## Déséquilibre de liaison



		Locus 1	
		A	a
Locus 2	b	6	0
	B	0	6

$$R^2 = 1$$

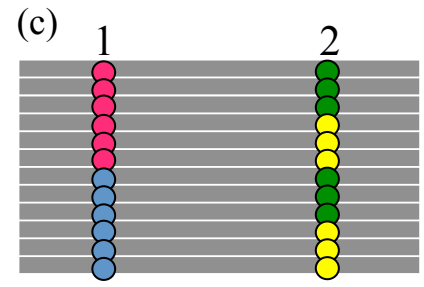
$$D' = 1$$



		Locus 1	
		A	a
Locus 2	b	6	3
	B	0	3

$$R^2 = 0.3$$

$$D' = 1$$



		Locus 1	
		A	a
Locus 2	b	3	3
	B	3	3

$$R^2 = 0$$

$$D' = 0$$

D représente l'écart entre les fréquences haplotypiques observées et attendues

$$D_{ab} = (\Pi_{AB} - \Pi_A \Pi_B)$$

$$R^2 = \frac{(D_{ab})^2}{\Pi_A \Pi_a \Pi_B \Pi_b}$$

$$D' = \frac{(D_{ab})^2}{\min(\Pi_A \Pi_b, \Pi_a \Pi_B)} \text{ si } D_{ab} < 0$$

$$D' = \frac{(D_{ab})^2}{\min(\Pi_A \Pi_B, \Pi_a \Pi_b)} \text{ si } D_{ab} > 0$$

**Figure 1.11:** Représentation schématique (a et b) du déséquilibre de liaison et (c) de l'équilibre de liaison entre deux loci. Le locus 1 contient les allèles A et a, et le locus 2 les allèles B et b avec une fréquence allélique de  $\Pi_A$ ,  $\Pi_a$ ,  $\Pi_B$  et  $\Pi_b$  et haplotypique de  $\Pi_{AB}$ ,  $\Pi_{Ab}$ ,  $\Pi_{Ba}$  et  $\Pi_{ba}$ . (d'après Flint-Garcia et Thornberry, 2003).

individus (Tableau 1.4), ce qui n'est souvent pas suffisant pour encadrer finement le QTL. Par ailleurs, ne sont détectables que les variations génétiques à l'état hétérozygote chez les parents (Figure 1.10), avec donc un maximum de 4 allèles en ségrégation dans des populations de type plein frère. Enfin, chez les plantes pérennes, le niveau de résolution du QTL dépend du nombre de recombinaisons, souvent limité, survenues en une seule génération car il est très difficile de créer des populations dérivées de plusieurs générations de croisements, en raison de la sensibilité de la vigne à la dépression de consanguinité et de la longueur de son cycle biologique. Pour les mêmes raisons, le clonage positionnel de gènes d'intérêt, qui nécessite le phénotypage et le génotypage de très grandes populations reste un travail coûteux en ressources chez la vigne et explique que malgré le nombre de projets en cours, peu de gènes ont été publiés par cette stratégie.

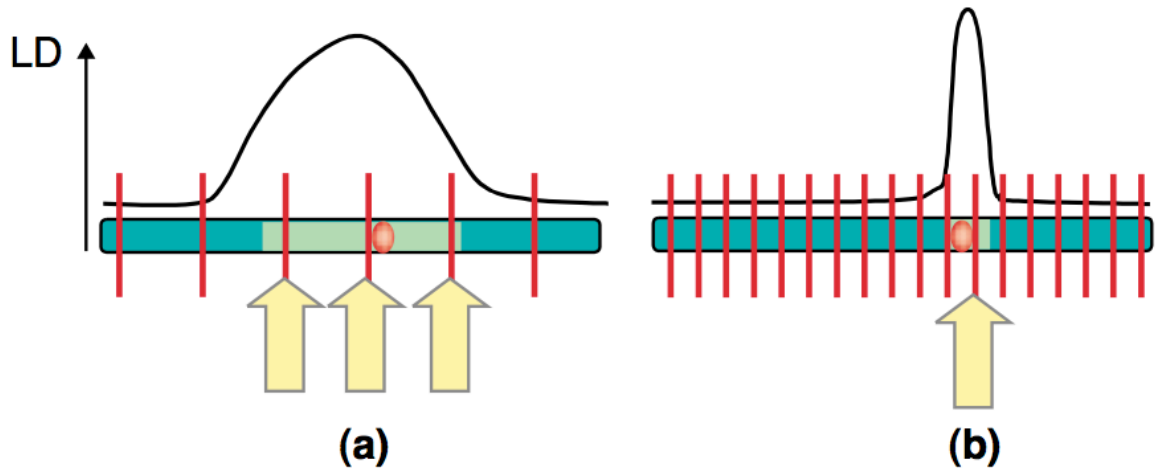
## 1.2.2. Génétique d'association

Chez l'Homme, pour pallier à l'impossibilité de mettre en place des dispositifs d'études QTL puissants, des stratégies de détection d'associations statistiques entre phénotypes et génotypes dans de larges populations ont été développées (Altshuler et al, 2008). La génétique d'association consiste, comme la recherche de QTL, à détecter des corrélations entre la variation d'un phénotype et des polymorphismes de marqueurs génétiques mais cette fois-ci, dans une population naturelle composée d'individus non apparentés et présentant donc une grande diversité génétique et phénotypique (Figure 1.10). Cette stratégie repose sur l'utilisation de la diversité génétique naturellement présente chez l'espèce considérée, ainsi que sur des structures chromosomiques ayant évoluées au cours du temps en accumulant un grand nombre de recombinaisons. Elle capture ainsi potentiellement un maximum de la diversité existante chez l'organisme étudié (Figure 1.10).

### 1.2.2.1. Déséquilibre de liaison et test d'association: généralités

La cartographie par génétique d'association s'appuie sur le Déséquilibre de Liaison (DL). Le DL est une association non aléatoire d'allèles à différents loci (Figure 1.11; Hartl et Clark, 1997). Le DL décrit une situation dans laquelle certaines combinaisons d'allèles ou de marqueurs génétiques dans une population sont plus fréquentes que ce que l'on attendrait d'une combinaison aléatoire de ces allèles en fonction de leur fréquence (Flint-Garcia et al, 2003). Différentes statistiques permettent d'estimer le DL entre deux loci.  $R^2$  et  $D'$  sont les plus couramment utilisées et sont présentées dans la Figure 1.11. La valeur de  $R^2$  est

LD	Fort (a)	Faible (b)
Résolution	faible	forte
Nombre de marqueurs	faible	élevé
Stratégie	WGAS	gène candidat



**Figure 1.12:** Relation entre l'étendu du LD et la résolution d'études de génétique d'association. En (a) le LD (zone bleu ciel) diminue doucement de part et d'autre du gène responsable du phénotype (ovale orange) et demande donc une densité de marqueurs faible (barres verticales rouges) le long du chromosome pour la mise en évidence d'associations marqueurs/phénotype (flèches jaunes). En (b) le LD diminue rapidement autour du gène d'intérêt et une densité plus importante de marqueur est requise pour identifier une association marqueur-phénotype (d'après Rafalski, 2002).



influencée par le taux de recombinaison et de mutation alors que la valeur de  $D'$  est seulement influencée par le taux de recombinaison (Flint-Garcia, et al, 2003). Etant donné que  $R^2$  est un meilleur indicateur de la nature de la corrélation entre le marqueur et le gène responsable de la variation phénotypique, il est plus couramment utilisé dans les études d'association (Abdallah et al, 2003).

Le DL peut être influencé par de nombreux facteurs (Rafalski et Morgante, 2004; Gupta et al, 2005; Slatkin, 2008), qui tendent soit à le diminuer (fort taux de recombinaison, fort taux de mutation, allogamie), soit à l'augmenter globalement (autogamie, histoire de la population: goulot d'étranglement, structure de la population, taille de la population: petite population en dérive génétique, faible taux de recombinaison) ou localement (sélection balancée ou purificatrice, réarrangements chromosomiques qui suppriment la recombinaison locale).

La génétique d'association reposant sur la détection de marqueurs en DL avec un gène responsable d'une variation phénotypique, plusieurs stratégies peuvent être envisagées en fonction de l'étendue du DL chez l'espèce d'intérêt (Figure 1.12; Rafalski, 2002). Si le DL est fort, une analyse sans *a priori* du génome entier (*Whole Genome Association Scan*: WGAS) peut être envisagée à l'aide d'une quantité raisonnable de marqueurs mais aboutira à une résolution faible pour l'identification du gène sous-jacent. Si le DL est faible, une approche WGAS ne pourra être réalisée qu'en utilisant de très nombreux marqueurs (parfois des millions), permettant en retour une résolution plus fine pour l'identification du gène responsable du phénotype. Le coût des méthodes de génotypage à très haut débit reste encore à ce jour prohibitif en cas de faible DL espéré chez une espèce. Les équipes se tournent alors préférentiellement vers l'utilisation de la génétique d'association pour la validation de gènes candidats dans l'intervalle d'un QTL, permettant de limiter le nombre de marqueurs utilisés (par exemple chez la vigne: Fournier-Level et al, 2009; Emanuelli et al, 2010).

Plusieurs études WGAS ont cependant été conduites avec succès chez des espèces autogames présentant donc un fort DL, comme par exemple chez *Arabidopsis* (DL de 10-500 kb; Kim et al, 2007) en recherchant des associations entre la variation de 107 caractères phénotypiques et 250 000 SNP à l'aide d'une puce Affimetrix sur 192 génotypes (Atwell et al, 2010), ou chez le riz (DL de 75-500 kb; Mather et al, 2007) avec 14 caractères d'intérêt agronomique et environ 3,6 millions de SNP issus du séquençage de 517 variétés. Un nouveau type de population, utilisé en WGAS a très récemment été développé avec succès chez le maïs pour l'identification des bases génétique de la floraison (Buckler et al, 2009).

Ces populations appelées NAM, pour Nested Association Mapping, (Yu et al, 2008) constituent un outil performant pour valider rapidement et en grande quantité des gènes candidats responsables des caractères observés, et pour caractériser finement des QTL chez les plantes. Cependant, la création de ces lignées demande plusieurs cycles de croisement et d'autofécondation et ne peut pas être envisagée chez une espèce pérenne comme la vigne.

#### 1.2.2.2. Déséquilibre de liaison et test d'association chez la vigne

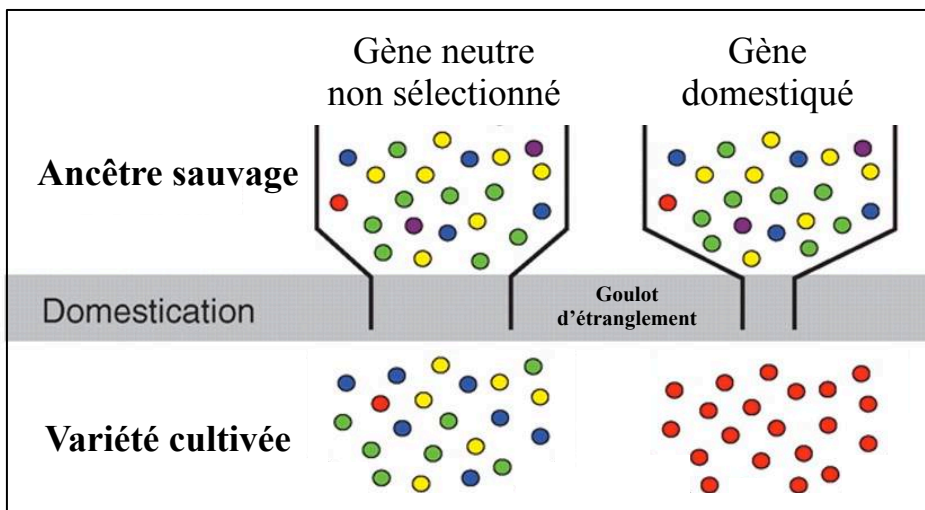
Plusieurs études ont permis de donner une première estimation du DL chez la vigne. Deux études, dans cinq régions correspondant toutes à des chromosomes différents et dans lesquelles avaient été détectés des QTL, ont été scannées à l'aide de 38 marqueurs SSR dans une core-collection de 141 de *V. vinifera* cultivées (Barnaud et al, 2006) et 85 *V. vinifera* sauvages (Barnaud et al, 2009). Un DL significatif jusqu'à 16,8 cM et 1,4 cM a été observé dans le compartiment cultivé et sauvage respectivement. Le goulot d'étranglement lié à la domestication de *V. vinifera* ainsi que la propagation par voie végétative des cultivars, est certainement à l'origine de cette différence. Cependant, l'estimation du DL au niveau nucléotidique d'un échantillon de la collection de génotypes cultivés a montré une diminution plus rapide du DL (This et al, 2007). Cette différence, aussi observée dans d'autres espèces, serait liée à la différence de mode d'évolution entre les deux types de marqueurs utilisés (Flint-Garcia et al, 2003; Stich et al, 2006). Le faible niveau du DL au niveau nucléotidique chez *V. vinifera* a été confirmé par trois autres études. La première porte sur l'analyse de 1500 SNP provenant de 200 loci répartis aléatoirement sur le génome sur 11 génotypes (9 cultivés et 2 sauvages) et montre une diminution du DL après 200 paires de bases (pb) (Lijavetzky et al, 2007). Les deux autres se basent respectivement sur des milliers de SNP identifiés par re-séquençage Illumina de 10 *V. vinifera* cultivées et 7 sauvages (Myles et al, 2010) et sur le génotypage de 950 *V. vinifera* cultivées et 59 sauvages à l'aide d'une puce SNP de 9K (Myles et al 2011). Elles montrent que le DL diminue rapidement après 3 kb chez la vigne (Myles et al, 2010; Myles et al, 2011).

La mise en œuvre d'une stratégie WGAS chez la vigne nécessiterait donc de génotyper des centaines de milliers de marqueurs SNP ce qui était inenvisageable au début de ma thèse. J'ai donc réalisé une étude d'association ciblée sur des gènes candidats et en particulier des gènes candidats co-localisant avec des QTL.

En effet, la génétique d'association a tout d'abord été utilisée en complément de détectations de QTL chez la vigne. La première étude publiée par Fournier-Level et al (2009) est la combinaison d'une étude de cartographie de QTL dans une descendance F1 (191 individus issus du croisement Syrah x Grenache) ségrégeant pour le contenu en anthocyanes de la baie, et d'une étude de génétique d'association dans une population naturelle (141 génotypes), pour tester les gènes candidats dans l'intervalle d'un QTL. Un QTL majeur expliquant jusqu'à 62% de la variation en anthocyanes de la baie a été identifié dans un intervalle contenant 68 gènes. Parmi ces gènes, se trouvait le cluster de 4 gènes de la famille *VvMybA* (*VvMybA1*, *VvMybA2*, *VvMybA3*, et *VvMybA4*). Le gène *VvMybA1* étant connu pour être impliqué dans la présence ou l'absence de pigmentation de la baie, la région contenant l'ensemble du cluster a donc été ciblée par la suite. Cinq polymorphismes (1 rétrotransposon, 3 SNP et un INDEL) dans 3 des 4 gènes *VvMybA* ont pu être identifiés par génétique d'association comme étant impliqués dans 84% de la variation en anthocyanes de la baie dans une population naturelle maximisant la variabilité phénotypique du contenu en anthocyane (Fournier-Level et al, 2009). Le second exemple, très récent (Emanuelli et al, 2010) a permis de mettre en évidence une association entre la variation en arôme muscaté de la baie dans une collection de 148 génotypes et 1 SNP non synonyme dans une 1-deoxy-D-xylulose-phosphate synthase (*VvDXS*) co-localisant avec un QTL pour ce même caractère (Battilana et al, 2009; Duchêne et al, 2009).

Une approche WGAS a également été testée très récemment chez la vigne (Myles et al, 2011). 950 *V. vinifera* cultivées ont été génotypés à l'aide de 5387 SNP provenant d'une puce SNP de 9K. Des associations entre des SNP et la variation du contenu en anthocyanes de la baie ont été recherchées et une région de 5 Mb sur le chromosome 2 contenant le même cluster de gènes *VvMybA* précédemment identifié par Fournier-Level et al (2009) a été révélée avec une forte sélection autour de ce locus chez les cultivars produisant des baies blanches. Ces résultats indiquent qu'une densité de marqueurs relativement faible semble suffisante pour identifier des associations phénotype/marqueur par la stratégie WGAS chez la vigne pour des caractères fortement sélectionnés donc présentant un forte DL (Fournier-Level et al, 2009). Par ailleurs, le faible nombre de générations de sélection chez la vigne et l'immortalisation des structures génétiques par bouturage sont certainement à l'origine de l'étendue du DL dans ces régions sélectionnées.

En utilisant des populations naturelles, les études de génétique d'association permettent de capturer un maximum de la diversité génétique présente chez l'organisme



**Figure 1.13:** Effet de la domestication sur la diversité génétique des plantes. Les ronds de couleur représentent les différents allèles d'un locus (d'après Yamasaki et al, 2007).

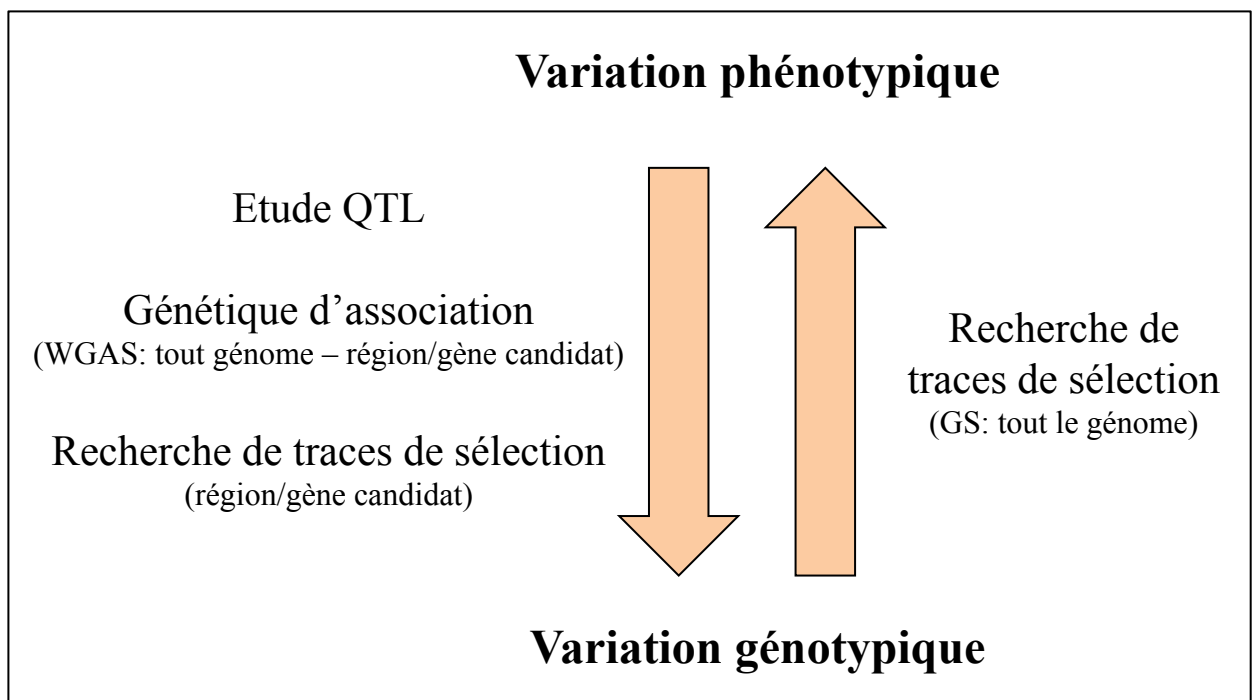
étudié. Il est alors théoriquement possible d'identifier l'intégralité des bases génétiques d'un caractère. Cette capacité à analyser un maximum de combinaisons alléliques rend la méthode très puissante (Yu et al, 2006; Zhu et al. 2008). Cependant, la puissance de détection d'une association dépend de la fréquence allélique. Des variants fonctionnels avec une faible fréquence ont peu d'influence sur la population entière et donc sont difficiles à détecter, même si ces faibles fréquences alléliques ont un effet important sur le phénotype. Dans ce cas, une approche de détection de QTL dans une population en ségrégation sera plus appropriée.

L'intérêt de travailler avec des populations naturelles est également de pouvoir affiner les résultats en testant les polymorphismes dans des fonds génétiques très diversifiés (Flint-Garcia et al, 2005), ce qui rend les résultats obtenus extrapolables à tous les génotypes existants. De plus, du fait de nombreuses recombinaisons dues à l'utilisation d'une population naturelle, lorsque l'on détecte une association, l'espérance que le polymorphisme causal de la variation soit dans un intervalle proche du marqueur est d'autant plus forte que le DL est faible, ce qui est le cas chez la vigne (Myles et al, 2010). On peut donc espérer obtenir un très bon niveau de résolution jusqu'au niveau du polymorphisme causal de la variation. La génétique d'association représente donc un outil de choix dans le raffinement des études de QTL lorsque le clonage positionnel est techniquement irréalisable du fait de la biologie de l'espèce considérée (Neale et Savolainen, 2004).

### **1.2.3. Recherche de traces de sélection**

#### **1.2.3.1. Généralités**

La domestication des plantes par l'Homme a entraîné une perte de la diversité génétique chez les variétés cultivées par rapport à la diversité génétique d'origine, présent chez les sauvages (Figure 1.13; Ralalski et Morgante, 2004; Yamasaki et al, 2007; Hamblin et al, 2011). Cette diminution a porté, d'une part sur l'ensemble du génome à cause du goulot d'étranglement lié au choix d'un petit nombre d'individus, (Tanksley et McCouch, 1997) et d'autre part sur des loci d'intérêts agronomiques et leur voisinage immédiat par effet auto-stop, due de la sélection d'allèles favorables à ces gènes (Rafalski et Morgante, 2004; Tenaillon et al, 2004). Une autre conséquence de la sélection est l'augmentation du déséquilibre de liaison autour des gènes sélectionnés par rapport à ce qui est observé dans le reste du génome (Wang et al, 1999; Palaisa et al, 2004). Bien sûr, l'étendue du DL dans les régions sous sélection dépend de la sévérité du goulot d'étranglement, du taux de recombinaison et de la distance physique entre les loci (Kim and Stephan, 2002). La théorie

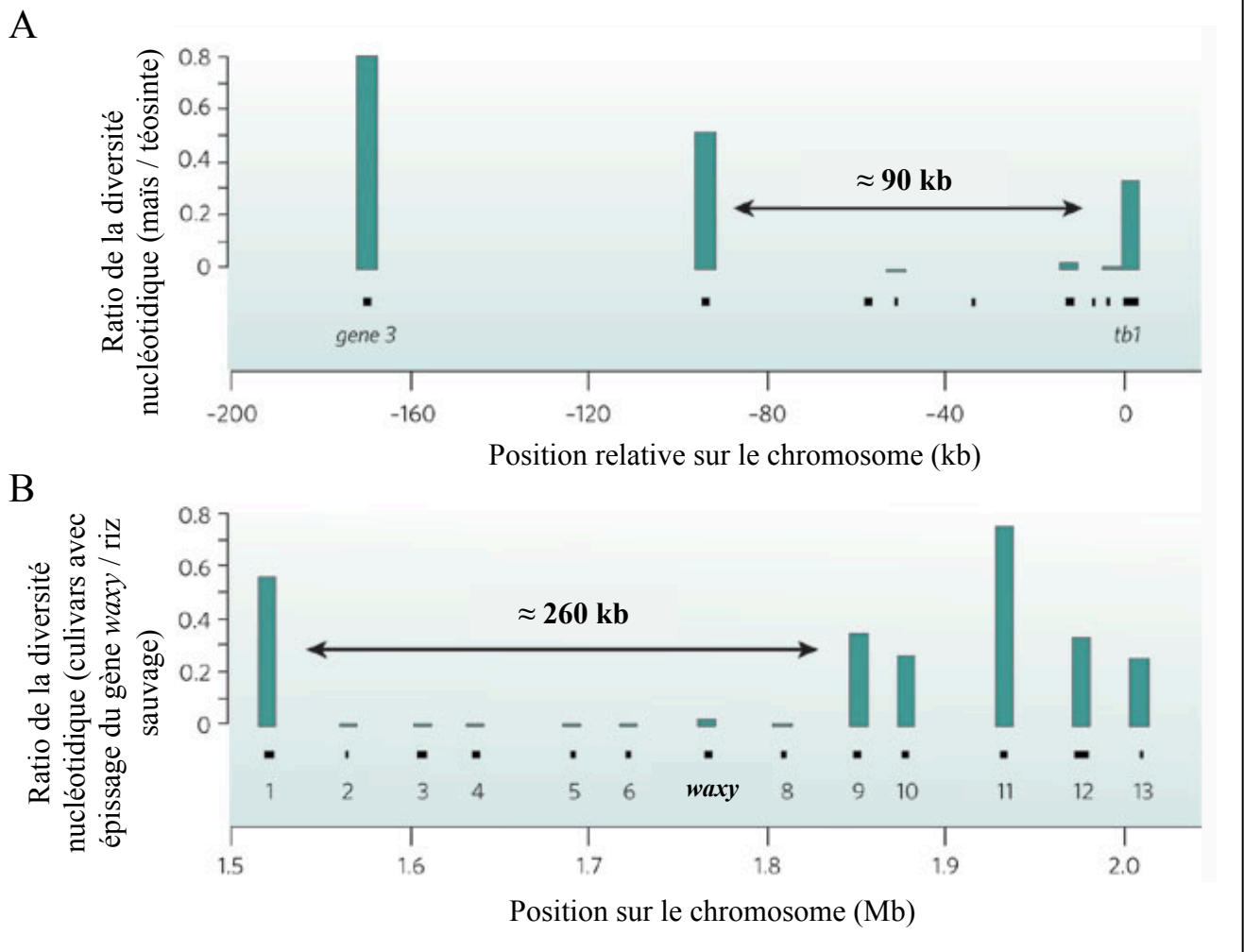


**Figure 1.14:** Différentes stratégies utilisées pour l'étude du déterminisme génétique de caractères quantitatifs.

neutraliste de l'évolution (Kimura, 1968) propose que la plupart des loci sont sélectivement neutres, c'est à dire qu'ils ne sont pas soumis à la sélection mais à la dérive et évoluent de manière aléatoire. Le modèle d'évolution basé sur cette théorie suppose en outre un échantillonnage aléatoire des gamètes et une population effective de très grande taille. Plusieurs tests permettent de détecter un écart au modèle d'évolution neutre, et ainsi de rechercher des traces de sélection en analysant la diversité de séquences d'ADN. Le test le plus couramment utilisé est le test du D de Tajima (Tajima, 1989). Ce test dépend de deux paramètres:  $\theta$  le nombre de sites en ségrégation (Watterson, 1975) et  $\pi$  le nombre moyen de différences nucléotidiques par site entre paires de séquences prises au hasard (Tajima, 1983). Sous le modèle d'évolution neutre, le D de Tajima vaut zéro. Une valeur négative indique un excès d'allèles rares, témoignant soit d'une sélection purificatrice (sélection d'un allèle favorable aboutissant à une diminution de la diversité à ce locus), soit d'une expansion géographique. Par opposition, une valeur positive suggère soit une sélection balancée (sélection de plusieurs allèles à un locus augmentant la diversité génétique), soit une réduction de l'effectif démographique, soit une mauvaise prise en compte de la structure de l'échantillon (Wright, 2005). Il est donc important de comparer les valeurs de ce test au locus d'intérêt avec ceux de gènes neutres afin de faire la distinction entre les effets dus à la sélection et ceux dus aux événements démographiques et d'analyser préalablement soigneusement la structure génétique de l'échantillon.

### 1.2.3.2. Exemples d'identification de traces de sélection

Les premières recherches de gènes sélectionnés pendant la domestication ont porté sur des analyses de QTL pour des caractères importants dans la domestication, dans des populations en ségrégations issues de croisements entre des variétés cultivées et des génotypes sauvages. De nombreux QTL ont ainsi été cartographiés (Paterson, 2002; Tanksley 2004; Gross et Olsen, 2010 pour revues). Cependant, beaucoup de ces études n'ont pas permis d'identifier les gènes sous-jacents aux QTL. Une autre approche sans *a priori* (*Genomic selection*: GS) (Wright et al, 2005; Figure 1.14) consiste à étudier le polymorphisme de séquences ADN le long des chromosomes, et de rechercher des régions génomiques présentant des traces de sélection dans un échantillon représentatif de la diversité génétique de variétés cultivées et de leurs progéniteurs potentiels sauvages. Ces traces se présentent sous la forme d'une différence de diversité nucléotidique entre génotypes sauvages et cultivés et d'un D de Tajima significativement divergent de l'hypothèse de sélection neutre. Afin d'identifier



**Figure 1.15:** Exemples de traces de sélection dans la région (A) du gène *tb1* chez le maïs et (B) du locus *waxy* chez le riz. Les traces de sélection correspondent à une diminution de la diversité nucléotidique chez les génotypes cultivés par rapport aux sauvages, au niveau du gène cible et des gènes environnants. Les rectangles noirs représentent les différents gènes présents dans les régions étudiées (Purugganan et Fuller, 2009).



à quels phénotypes sont reliées les régions portant des signatures de sélection, des associations entre les polymorphismes de séquences présents dans ces régions et la variation de caractères phénotypiques sont recherchées. Les régions portant des traces de sélection sont également confrontées aux résultats obtenus dans le cadre d'étude de détection de QTL. Ce type de stratégie a permis d'identifier des zones du génome ayant subi une forte sélection au cours de la domestication chez le maïs (2 à 4% des gènes; Wright et al, 2005; Yamasaki et al, 2007), le sorgo (environ 15% des gènes; Casa et al, 2005; Hamblin et al, 2006), mais également le tournesol (environ 7% des gènes; Chapman et al, 2008). Ces études sont validées par le fait que de nombreuses régions présentant des traces de sélection co-localisent avec des QTL déjà identifiés comme intervenant dans des différences morphologiques entre sauvages et cultivées (Hamblin et al, 2006; Yamasaki et al, 2007; Chapman et al, 2008). Mais elles ont également permis de mettre en évidence de nouveaux gènes impliqués par exemple dans la croissance de la plante, la maturité et la qualité nutritionnelle du fruit et de la graine (Wright et al, 2005; Chapman et al, 2008). La recherche de traces de sélection sans *a priori* demande des centaines de milliers de marqueurs et une bonne connaissance du compartiment cultivé et sauvage, dont dépendent la nature des traces identifiées et le taux de faux positifs. Enfin, comme pour les études de génétique d'association, cette technique reposant sur le DL créé par la sélection, un fort DL chez l'espèce d'intérêt conduira à une résolution faible du gène sous-jacent et inversement.

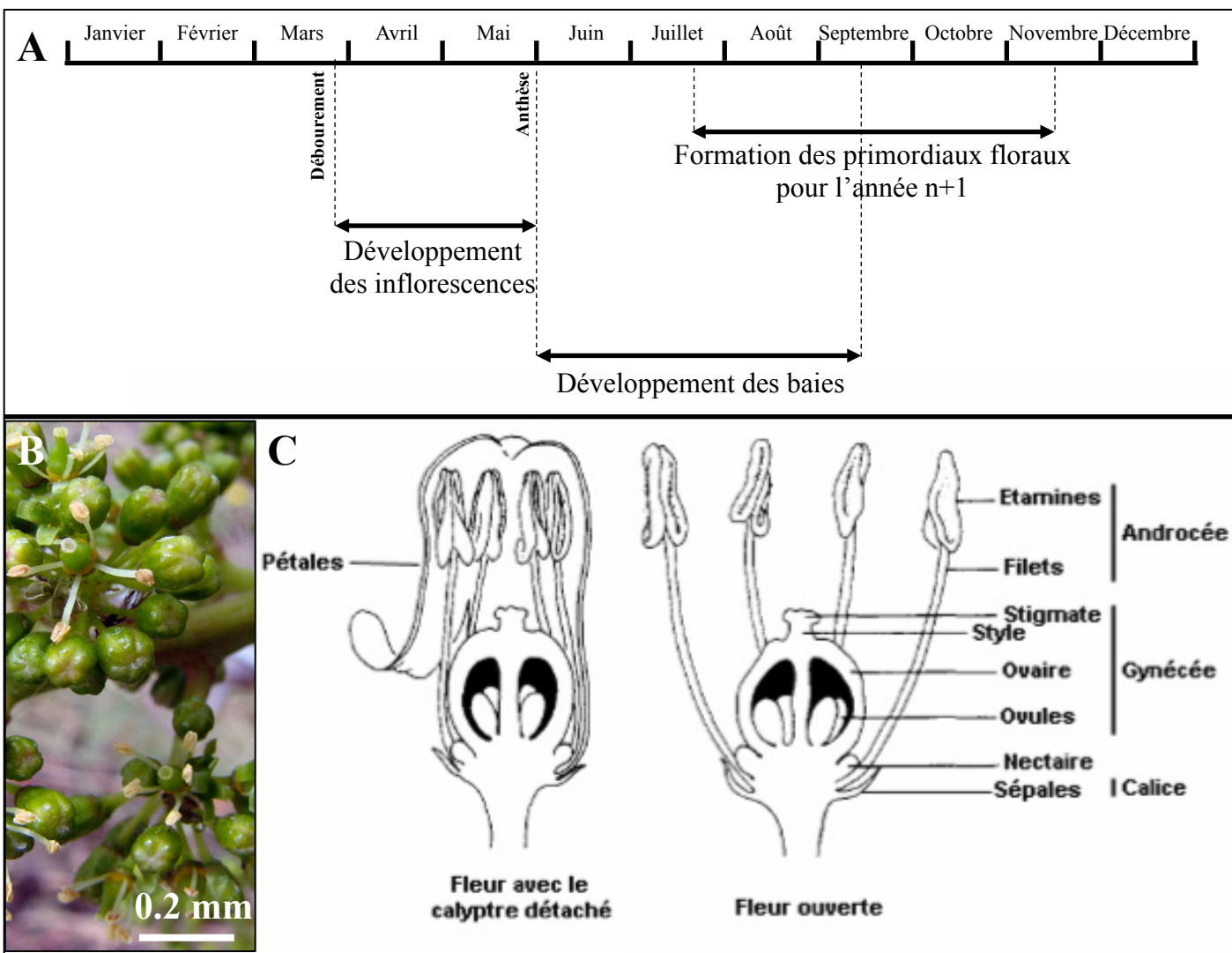
La recherche de traces de sélection peut également être réalisée à plus petite échelle en complément à des détections de QTL et devrait en principe aider à trier des gènes candidats dans les régions pointées ou bien pour mieux comprendre l'histoire évolutive de gènes déjà identifiés (Nesbitt et Tanksley, 2002; Camus-Kulandaivelu et al, 2008). En effet, la plupart des QTL identifiés chez les plantes cultivées étant responsables de la variation de caractères agronomiques, il est plus que probable qu'ils aient été soumis à la sélection et que la région contenant ces QTL continue à porter les traces de cette sélection (Hamblin et al, 2011). Des traces de sélection ont ainsi été mises en évidence dans des gènes impliqués dans la transition de la téosine au maïs, comme le gène *tb1* impliqué dans la formation de branches secondaires (Clark et al 2004 et 2006) et le gène *Y1* impliqué dans la couleur de l'albumen (Palaisa et al, 2003 et 2004). Dans le cas de *tb1*, la trace de sélection correspond à une zone de 90 kb en amont de la partie codante du gène montrant une diminution de la diversité génétique chez le maïs pas rapport à la téosine (Figure 1.15) et pour *Y1* à une région d'environ 600 kb. D'autres exemples sont disponibles comme pour le riz dont le locus *waxy*, associé à une faible quantité



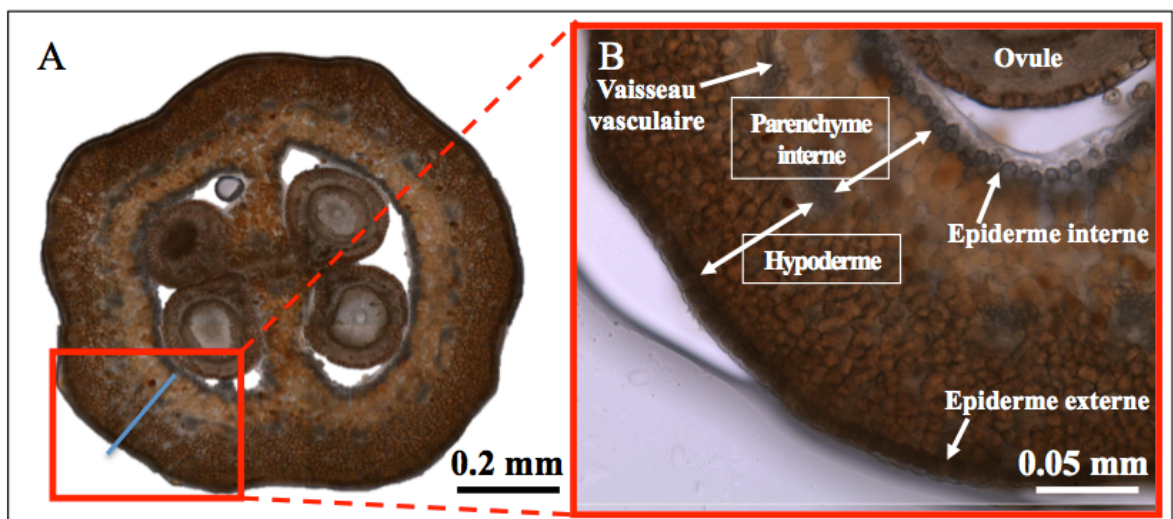
d'amylose par épissage du gène dans les cultivars du Nord-Est de l'Asie, montre une trace de sélection sur 260 kb (Figure 1.15; Olsen et al, 2006).

Chez la vigne, une étude très récente (présentée précédemment; Myles et al, 2011) basée sur le génotypage de 950 *Vitis vinifera* cultivées et 59 sauvages à l'aide de 5387 SNP répartis le long du génome, a permis d'identifier une région de 5 Mb sur le chromosome 17 portant une forte trace de sélection (une diversité haplotypique beaucoup moins importante chez les *V. vinifera* cultivées que chez les sauvages). Cette région n'a cependant pas encore été reliée à la variation d'un caractère morphologique.

Il est important de noter que la distinction entre les traces dues à la domestication de l'espèce et celles dues à la sélection après la domestication sont difficilement distinguables même si l'on compare le compartiment sauvage avec le cultivé. Il est fort probable que les gènes identifiés dans ces études ont été ciblés soit pendant la domestication, soit pendant les étapes de sélection qui ont suivi la domestication, soit pendant les deux. Enfin, l'identification de gènes sous sélection naturelle n'est pas à exclure même si les génotypes étudiés proviennent généralement de régions géographiques très différentes dans les deux compartiments considérés (Atwell et al, 2010). La recherche de traces de sélection est donc une approche complémentaire aux deux autres, pour identifier et caractériser les gènes importants pour la variation des caractères agronomiques ou l'adaptation des plantes à leur environnement, préliminaire indispensable à la sélection variétale moderne.



**Figure 1.16:** La fleur de vigne. (A) Cycle de développement des inflorescences et baies. Photographie (B) et schémas (C) de la fleur hermaphrodite de vigne et de son inflorescence à l'anthèse.



**Figure 1.17:** Tissus de l'ovaire. (A) Tissus de l'ovaire de vigne. Coupe transversale d'un ovaire entièrement formé avant fécondation. (B) Agrandissement des différents tissus composant le péricarpe de l'ovaire.

## 1.3. Bases physiologiques et génétiques de la variabilité des fruits charnus

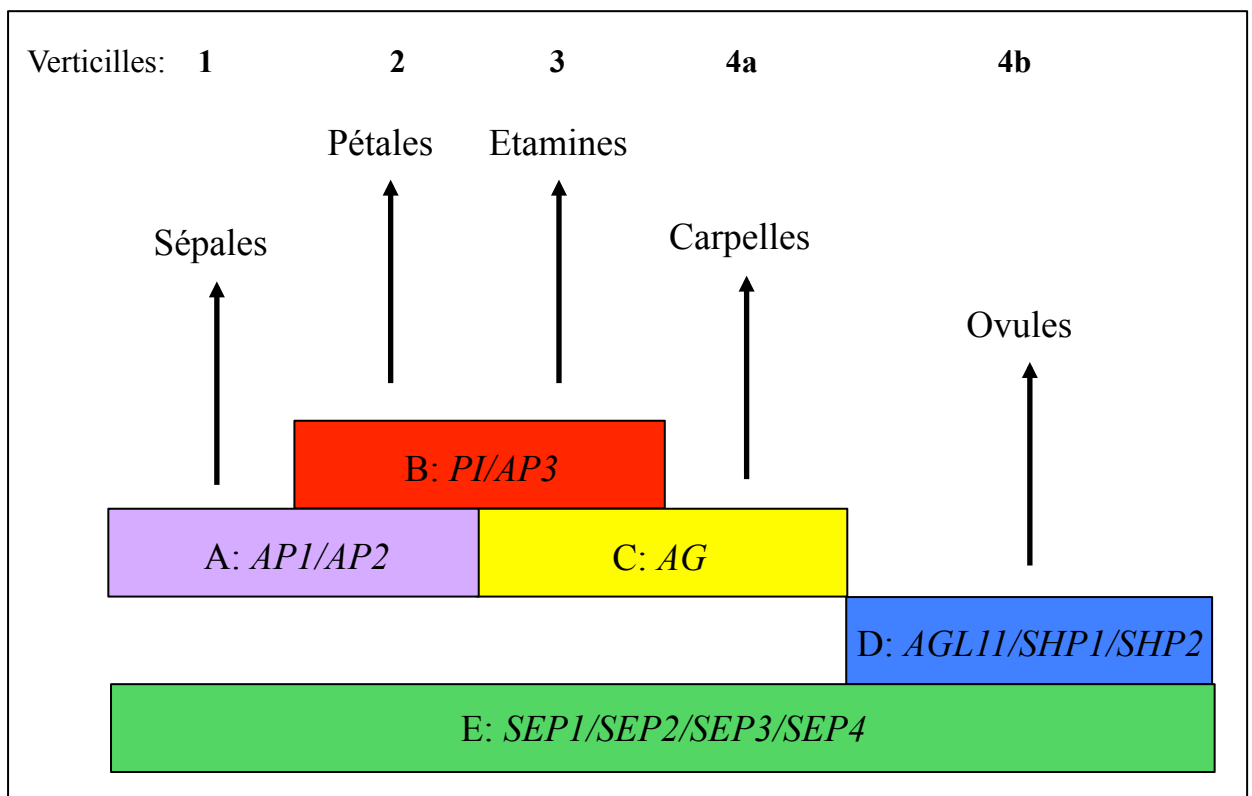
Cette partie décrit l'état des connaissances sur le développement de la baie de vigne plus particulièrement sous l'angle de la variabilité de la taille du fruit. Ces connaissances seront mises en parallèle de ce qui est connu chez les plantes modèles *Arabidopsis* pour le développement floral et la tomate pour le développement du fruit.

### 1.3.1. Développement floral chez la vigne

#### 1.3.1.1. Structure des fleurs de vigne

Les primordia inflorescentiels, qui au départ sont non distinguables des primordia des vrilles, se forment dans les bourgeons latents et leur différenciation se poursuit jusqu'à l'entrée en dormance des bourgeons (Carolus, 1971). Dès le débourrement, au printemps de l'année suivante (Figure 1.16-A; Carmona et al; 2008), le développement des inflorescences reprend et les différentes pièces florales se différencient. Au stade final du développement, la fleur de vigne hermaphrodite est composée de quatre verticilles (Figure 1.16-B et C):

- le calice, verticille le plus externe qui comprend cinq sépales soudés entre eux;
- la corolle, constituée de cinq pétales qui alternent avec les sépales et qui sont soudés entre eux, formant un capuchon appelé calypstre. Lors de la floraison, la corolle s'ouvre généralement par la base et le capuchon ainsi libéré est rejeté (Figure 1.16-C);
- l'androcée, formé par cinq étamines opposées aux pétales. Leur filet est long et leurs anthères présentent deux loges à déhiscence longitudinale introrse;
- le gynécée, verticille le plus interne, composé d'un style toujours court, d'un stigmate et d'un ovaire supère non adhérent formé de deux carpelles soudés au niveau de la columelle, renfermant chacun deux ovules. L'ovaire est formé de plusieurs tissus (Figure 1.17). En partant de l'extérieur vers l'intérieur, on trouve l'épiderme externe qui est constitué d'une seule couche cellulaire. L'hypoderme séparé du parenchyme interne par le système vasculaire et l'épiderme interne constitué d'une seule couche cellulaire délimitant les loges carpellaires contenant les ovules (Park, 1995). La vascularisation de l'ovaire apparaît au moment de la méiose, environ 3 semaines avant la floraison dans les parois carpellaires, puis dans la columelle (Benharbit El Alami, 1995).



**Figure 1.18:** Modèle ABCDE indiquant les différents domaines d'activité des gènes des classes A, B, C D et E dans le développement des pièces florales chez *Arabidopsis* (d'après Theissen et al, 2001).

**Tableau 1.5:** Liste des orthologues putatifs des gènes de classe A, B, C, D et E chez la vigne. L'expression de ces gènes a été testée dans les racines, feuilles, vrilles, fleurs juste avant l'anthèse et dans les baies à un stade précoce du développement (baie verte) et au stade véraison (d'après Diaz-Riquelme et al, 2009).

Nom du gène	Expression	Référence
<i>VvAP1</i>	vrille	Calonje et al, 2004; Diaz-riquelme et al, 2009
<i>VvPI</i>	fleur	Sreekantan et al, 2006; Fernandez et al, 2007; Diaz-riquelme et al, 2009
<i>VvAP3.1</i>	fleur	Poupin et al, 2007; Diaz-riquelme et al, 2009
<i>VvAP3.2</i>	fleur, baie verte et vérée	Poupin et al, 2007; Diaz-riquelme et al, 2009
<i>VvAG1</i>	fleur, baie verte et vérée	Boss et al, 2001; Diaz-riquelme et al, 2009
<i>VvAG2</i>	fleur, baie verte et vérée	Diaz-riquelme et al, 2009
<i>VvAGL11</i>	fleur, baie verte et vérée	Boss et al, 2002; Diaz-riquelme et al, 2009
<i>VvSEP1</i>	fleur	Boss et al, 2002; Diaz-riquelme et al, 2009
<i>VvSEP2</i>	fleur	Diaz-riquelme et al, 2009
<i>VvSEP3</i>	fleur	Boss et al, 2002; Diaz-riquelme et al, 2009
<i>VvSEP4</i>	fleur, baie verte et vérée	Diaz-riquelme et al, 2009

Chaque inflorescence contient de 100 à 1000 fleurs, chacune mesurant environ un millimètre (Figure 1.16-C).

### 1.3.1.2. Bases moléculaires du développement de la fleur de vigne

Le modèle synthétique de l'organogenèse florale a largement été décrit chez l'espèce modèle *Arabidopsis* et le paragraphe qui suit décrit l'état actuel des connaissances chez cette espèce en mettant en parallèle ce qui est connu chez la vigne.

La combinaison de cinq types de gènes MADS box codant pour les facteurs de transcription déterminent le devenir des différentes zones du méristème floral chez *Arabidopsis* (Figure 1.18; Schwarz-Sommer et al, 1990; Bowman et al, 1991; Coen et Meyerowitz, 1991; Weigel et Meyerowitz, 1994; Angenent et Colombo, 1996; Pélaz et al, 2000 et 2001; Jack, 2001; Theissen, 2001; Ditta et al, 2004; Theissen et Melzer, 2007). Selon ce modèle, connu sous le nom de modèle ABCDE, les gènes de classe A (*Apetala 1: AP1* et *Apetala 2: AP2*) spécifient à eux seuls la formation des sépales, ou combinés à ceux de la classe B (*Apetala 3: AP3* et *Pistillata: PI*) la formation des pétales. Les gènes de classe B à leur tour combinés à ceux de classe C (*Agamous: AG*) spécifient la formation des étamines, tandis que les gènes C seuls, déterminent le développement des carpelles. Dans ce modèle, les activités A et C sont mutuellement antagonistes. Lorsqu'un de ces gènes est muté, l'organogénèse florale est altérée et un verticille floral se transforme en un autre. L'activité D a été identifiée chez le Pétunia avec les gènes *Floral binding protein 7* et *11* (*FBP7* et *FBP11*) qui contrôlent la formation des ovules et les gènes correspondant chez *Arabidopsis* sont *Agamous-like 11* (*AGL11*) et *Shatterproof 1* et *2* (*SHP1* et *2*). Les quatre autres gènes de classe E (*Sepellata 1, 2, 3* et *4* ou *SEPI, 2, 3* et *4*) interviennent de manière redondante: les triples mutants E voient leurs pièces florales transformées en sépales, ce qui signifie que ces gènes sont nécessaires pour l'activité des gènes de classe B et C. Ainsi pour obtenir un sépale, l'activité A est suffisante, les interactions A+B+E, B+C+E et C+E sont respectivement nécessaires pour l'identité d'un pétale, d'une étamine et d'un carpelle. Les fonctions A, B, C, D, E sont limitées respectivement aux verticilles 1 et 2; 2 et 3; 3 et 4a; 4b; 2 à 4b (Figure 1.18).

D'autres gènes MADS box ou *Agamous-like* (*AGL*) ont été identifiés chez *Arabidopsis* avec des fonctions au cours du développement de la plante autres que la formation florale, telles que le gène *Fruitfull* (*FUL*) dont la mutation provoque des fruits indéhiscent et anormalement réduits (Gu et al, 1998; Ferrandiz et al, 1999). Ces résultats ont permis de



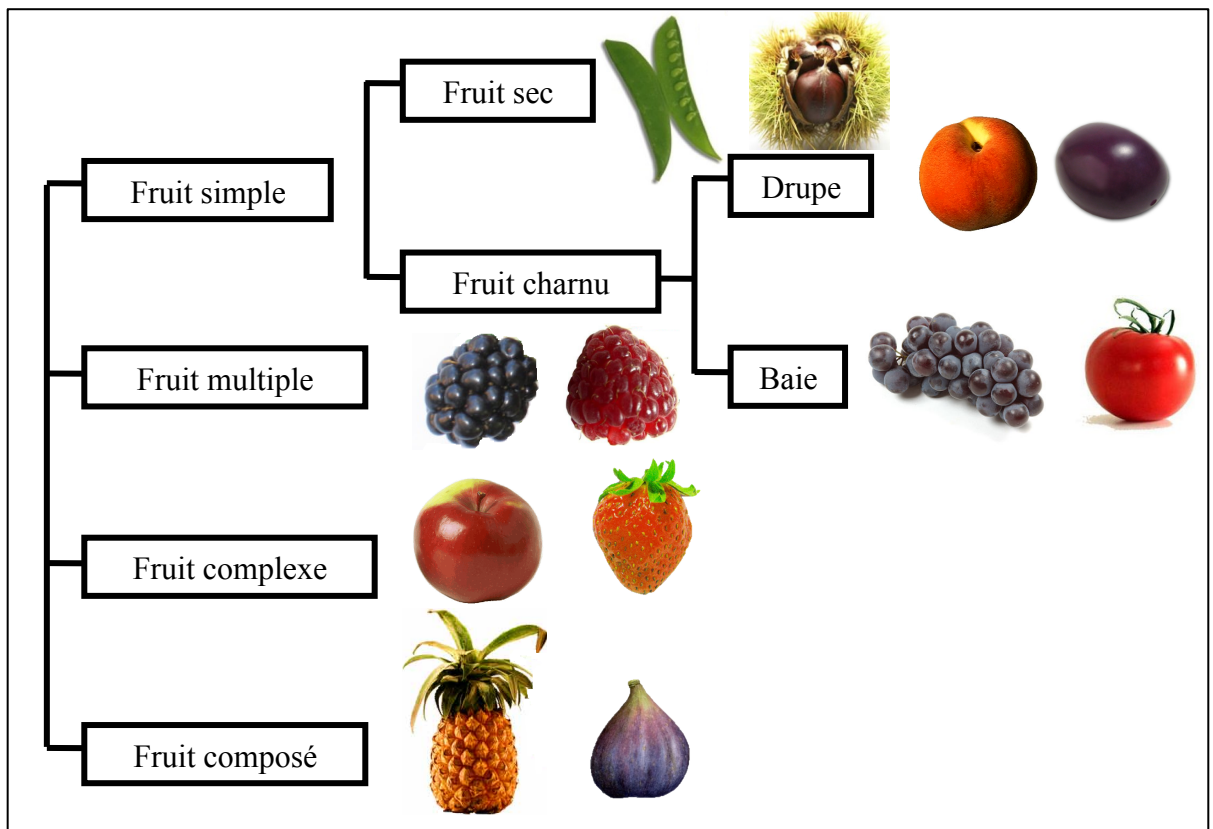


montrer que ce gène contrôle en partie la différenciation et la croissance cellulaire dans les valves (parois de l'ovaire). Ces résultats suggèrent que des gènes de type MADS-box sont d'importants facteurs contrôlant à la fois la différenciation et le développement des tissus de la fleur et du fruit.

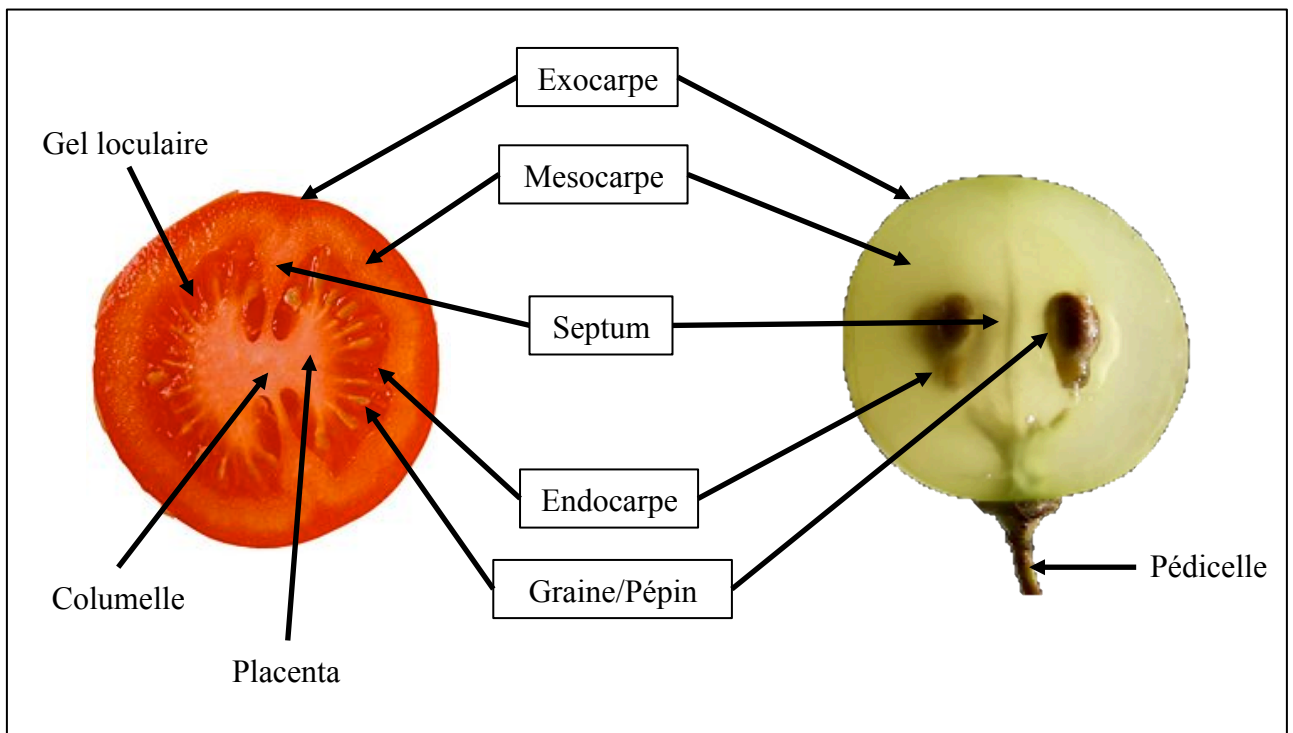
Chez la vigne, le rôle de plusieurs MADS box a pu être mis en évidence au cours de développement de la fleur et du fruit (Poupin et al, 2007; Chatelet et al, 2007). Des homologues putatifs (basés sur une similarité de séquences) des gènes de classe B, C, D et E ont été identifiés (Tableau 1.5). Il existe trois gènes chez la vigne avec une fonction similaire à celle des gènes de classe B: *VvPI*, *VvAP3.1*, *VvAP3.2* (Sreekantan et al, 2006; Poupin et al, 2007). Plusieurs homologues putatifs ont également été identifiés pour les gènes des classes C, D et E: *VvAG1*, *VvAG2*, *VvAGL11*, *VvSEP1*, *VvSEP2*, *VvSEP3* et *VvSEP4* (Boss et al, 2001; Boss et al, 2002; Fernandez et al, 2007; Diaz-Riquelme et al, 2009). Le gène *VvAPI* dont la séquence présente une forte similarité avec celle du gène *API* de classe A chez *Arabidopsis* semble cependant jouer un rôle différent chez la vigne puisqu'il ne s'exprime que dans les vrilles alors que tous les autres gènes supposés être impliqués dans l'identité florale chez la vigne s'expriment bien dans les fleurs (Tableau 1.5; Diaz-Riquelme et al, 2009). Enfin aucun gène homologue aux gènes *SHP1* et *SHP2* n'a été identifié chez la vigne. Cependant, d'autres gènes de type MADS box dont *VvAGL6*, *VvFUL*, *VvFUL-L*, *VvTM6*, *VvAGL17.1*, *VvAGL15.1*, *VvAGL15.2*, *VvBS1* et *VvBS2* sont exprimés dans les fleurs et/ou les baies (Diaz-Riquelme et al, 2009). Il serait maintenant nécessaire d'étudier l'effet de la suppression de l'expression de ces gènes pour élucider définitivement leur rôle, puisque l'exemple de *VvAPI* montre bien que l'homologie de séquence ne suffit pas à prédire la fonction exacte du gène. Cependant, ce type de test fonctionnel est beaucoup plus long à finaliser chez la vigne que chez *Arabidopsis*. Une autre possibilité pourrait être de réaliser des études de complémentation de mutants d'*Arabidopsis* avec des gènes de vigne, comme cela a déjà été réalisé pour le gène *VvPI*, confirmant que ce gène est l'orthologue fonctionnel du gène *PI* d'*Arabidopsis* (Poupin et al, 2007).

### 1.3.2. Développement des fruits charnus

Le mot fruit vient du latin « fructus » qui signifie produire. Chez les angiospermes, le fruit est le produit de la fécondation d'une fleur ou d'une inflorescence et est constitué d'une enveloppe qui renferme la/les graine(s) (Coombe, 1976). Le développement du fruit est un



**Figure 1.19:** Classification simplifiée des fruits.



**Figure 1.20:** Anatomie de la baie (A) de vigne et (B) de la tomate.

processus essentiel de la reproduction des angiospermes car le fruit mûr protège, nourrit et constitue l'intermédiaire facilitant la dissémination des graines. Au cours de l'évolution, les espèces végétales ont développé diverses stratégies pour la dispersion des graines afin d'assurer la survie de leurs descendances (Coombe, 1976) aboutissant à une incroyable variabilité de structure, de forme, de couleur et de taille des fruits.

Les fruits peuvent être classés en quatre catégories (Figure 1.19; Coombe, 1976): fruit simple issu d'un ovaire d'une seule fleur, fruit multiple issu de plusieurs ovaires d'une même fleur, fruit complexe issu d'un ou plusieurs ovaire(s) et autres pièces florales d'une seule fleur, et fruit composé issu de plusieurs ovaires et autres pièces florales provenant de plusieurs fleurs. Les fruits peuvent ensuite être classés en deux sous-catégories: fruits secs et fruits charnus, selon que le péricarpe soit lignifié ou chargé d'eau à maturité, respectivement (Figure 1.19). Enfin, on peut encore classer les fruits charnus en drupes ou baies selon que l'endocarpe soit lignifié (forme un « noyau » entourant la graine) ou charnu, respectivement. Comme la tomate, le fruit de la vigne est un fruit simple puisqu'il provient de la fécondation d'un ovaire d'une seule fleur et c'est une baie car il est composé d'un péricarpe entièrement juteux à maturité (Esau, 1965). Par ailleurs, le raisin est classé parmi les fruits non-climactériques, par opposition aux fruits climactériques tels que la tomate, l'avocat, la poire ou encore la banane qui présentent un pic d'éthylène marquant le début de la maturation du fruit. Cependant, il faut noter qu'une production limitée d'éthylène a été signalée chez la baie de vigne (Alleweldt et Koch, 1977) et que des travaux ont montré que l'éthylène pouvait jouer un rôle dans la signalétique du déclenchement de la maturation de la baie (Chervin et al, 2004; Tesnière et al, 2004).

### 1.3.2.1. Anatomie de la baie de vigne et de tomate

La baie de vigne, tout comme celle de la tomate (Gillaspy et al, 1993) est constituée de trois tissus formant le péricarpe (Figure 1.20; Coombe, 1987; Park, 1995; Hardie *et al*, 1996; Ollat et al, 2002):

- l'épicarpe (ou exocarpe) riche en substances colorantes, aromatiques et tanins constitue la pellicule. Il dérive des cellules de l'épiderme externe et d'une partie des cellules de l'hypoderme de l'ovaire. L'épiderme est le tissu extérieur de la baie formé par une seule couche de cellules régulières à parois minces, de forme isodiamétrique couvertes par une épaisse cuticule et par de la cire. L'hypoderme est le tissu sous-épidermique qui est anatomiquement et physiologiquement différent des tissus plus



internes (Esau, 1965). Au début du développement de la baie, une à deux assises cellulaires hypodermiques sont visibles, pour comprendre à maturité en moyenne une dizaine de couches cellulaires ;

- le mésocarpe est composé de cellules parenchymateuses qui constituent la pulpe et sont caractérisées par leur grande taille, une paroi fine et une vacuole qui occupe la majeure partie du volume cellulaire. La partie du mésocarpe en contact avec la pellicule est appelée mésocarpe externe et dérive de la partie interne de l'hypoderme de l'ovaire, par opposition au mésocarpe interne situé du côté de l'endocarpe et qui dérive du parenchyme interne de l'ovaire (Figure 1.17 et 1.20). Les cellules du mésocarpe externe s'allongent tandis que les cellules du mésocarpe interne conservent des formes plus ou moins sphériques et peuvent atteindre 400 µm de diamètre, créant ainsi un déséquilibre entre l'épaisseur de mésocarpe externe et interne.
- l'endocarpe correspond au tissu le plus profond du péricarpe. Il forme les loges carpellaires au nombre de deux chez la vigne et variant de deux à plus de 20 chez la tomate (Cheniclet et al, 2005). Ces loges contiennent en tout, un maximum de quatre pépins chez la vigne et plusieurs dizaines de graines chez la tomate. L'endocarpe dérive de l'épiderme interne de l'ovaire et est formé de 2 à 3 couches cellulaires appelées hypoderme interne et de l'épiderme interne. L'endocarpe est constitué de cellules plates de petites tailles à cuticule épaisse et est souvent décrit comme faisant partie de la pulpe (Pratt, 1971).

La vascularisation de la baie dérive de celle de l'ovaire, même si au cours du développement, les faisceaux vasculaires périphériques se retrouvent excentrés vers la surface de la baie. La vascularisation de la baie est constituée d'un réseau de petits faisceaux périphériques alimentant le péricarpe et d'un faisceau central plus large alimentant les pépins.

Parallèlement aux cellules du mésocarpe, les cellules du septum se multiplient pour coloniser l'espace vide entre les loges carpellaires chez les deux espèces. Les cellules du septum sont irrégulières en taille et forme et représentent avec le mésocarpe les tissus majeurs de stockage des sucres, et participent à la pulpe (Harris et al, 1968).

Enfin, trois autres tissus spécifiques constituent la tomate: le placenta qui est attaché aux graines et permet leur développement, la columelle qui constitue la partie centrale du fruit et participe également à la chair de la tomate, et le gel loculaire qui entoure les graines. Bien que les tissus constituant le péricarpe, à l'origine de la pulpe chez les baies de vigne, soient conservés entre ces deux espèces, chez la tomate, d'autres tissus non présents chez la vigne sont à l'origine de la chair du fruit.



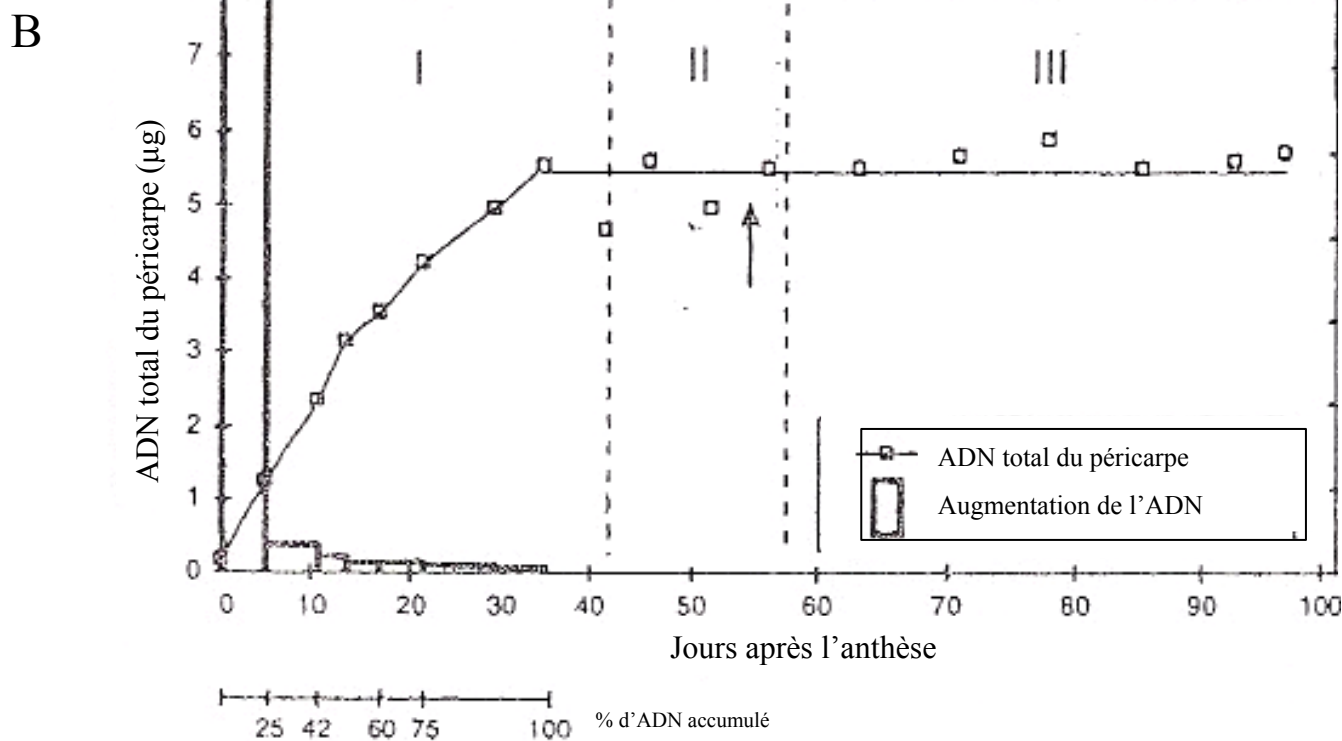
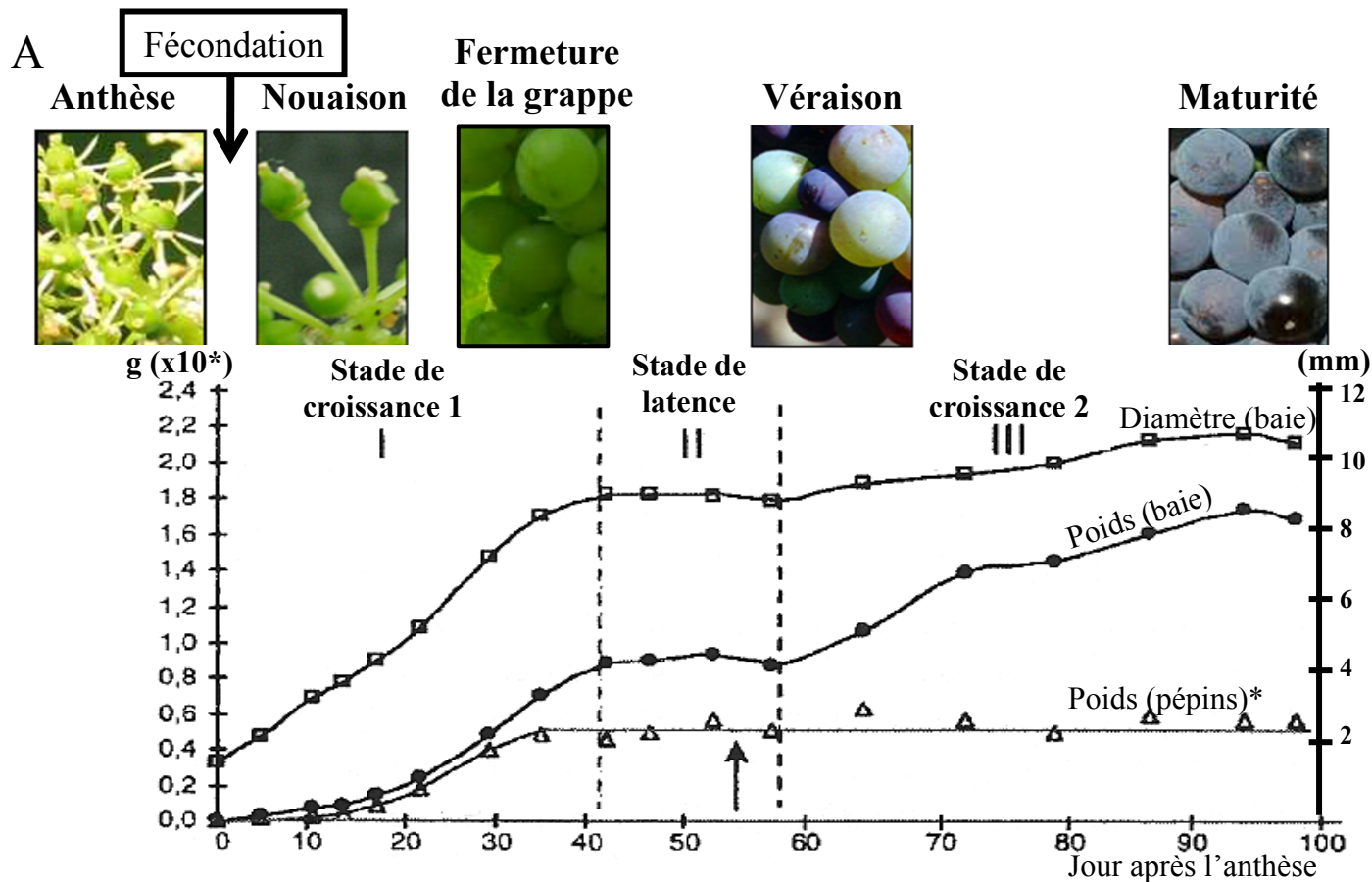
**Figure 1.21:** Photo d'une grappe de raisins contenant des baies millerandées distinguables par leur petite taille.

### 1.3.2.2. Développement de la baie de vigne

Chez la vigne cultivée, la fécondation a lieu en général 24 heures après l'anthèse (période pendant laquelle la fleur est complètement ouverte et fonctionnelle). Les quatre ovules présents sont rarement tous fécondés et on trouve alors en moyenne deux pépins par baie (Ollat et al, 2002). De manière générale, toutes les fleurs d'une inflorescence ne se transforment pas non plus en baie. Deux des anomalies de fécondation que l'on rencontre fréquemment chez la vigne et qui peuvent expliquer ce phénomène sont la coulure et le millerandage. Dans le premier cas, l'ovaire ne se transforme pas en fruit (Bessis et Fournioux, 1992) et dans le second cas, de petits fruits sans pépin (pathénocarpiques) se développent sans qu'il n'y ait eu de fécondation des ovules (Figure 1.21; Coombe, 1973). Après fécondation des fleurs, la paroi de l'ovaire se développe pour donner le péricarpe (Esau, 1965) composé de trois tissus: l'épicarpe, le mésocarpe et l'endocarpe. Un ovaire fécondé qui commence à se développer est dit « noué ». Le terme décrivant ce phénomène au niveau de la grappe est nouaison. Le pourcentage de nouaison est très variable en fonction des génotypes, des conditions environnementales et de l'alimentation de la vigne (Ebadi et al, 1995).

Le développement de la baie de vigne a fait l'objet de nombreuses études chez des variétés cultivées dont les résultats ont régulièrement fait l'objet d'articles et revues détaillées (Coombe 1960 et 1992; Ojeda et al, 1999 ; Ollat et al, 2002). La croissance des baies de vigne se caractérise par une évolution de poids de la baie en double sigmoïde (Figure 1.22), comme celle de la plupart des fruits charnus (Coombe, 1976). Selon les auteurs, elle est divisée en deux, trois, ou quatre stades de développement. La description en trois stades de développement (Ojeda et al, 1999) sera présentée ci-après (Figure 1.22):

- le stade I dure environ 40 jours à partir de la nouaison. Au cours de cette phase, la croissance des baies est liée à la fois à des phénomènes de division et de grandissement cellulaire, avec un pic des mitoses les cinq premiers jours suivants l'anthèse (Ojeda et al, 1999). Après ce pic, les divisions cessent rapidement pour se limiter à quelques assises cellulaires externes, les autres cellules du péricarpe étant déjà engagées dans une phase de grandissement cellulaire. La croissance de la baie résulte alors du grandissement cellulaire caractérisé par une forte accumulation d'eau et d'acides organiques dans la vacuole des cellules, notamment d'acides malique et tartrique (Schlosser et al, 2008). Ces acides sont synthétisés par les cellules de la baie, mais il a également été montré qu'une partie pouvait être importée (Glad et al, 1992;



**Figure 1.22:** Différents stades du développement de la baie de vigne. **(A)** Evolution du poids et du diamètre de la baie, ainsi que du poids des pépins pendant le développement de la baie. **(B)** Evolution de la quantité d'ADN au cours du développement de la baie chez le cultivar Syrah. La flèche représente le stade véraison (d'après Ojeda et al, 1999 et Park, 1995).

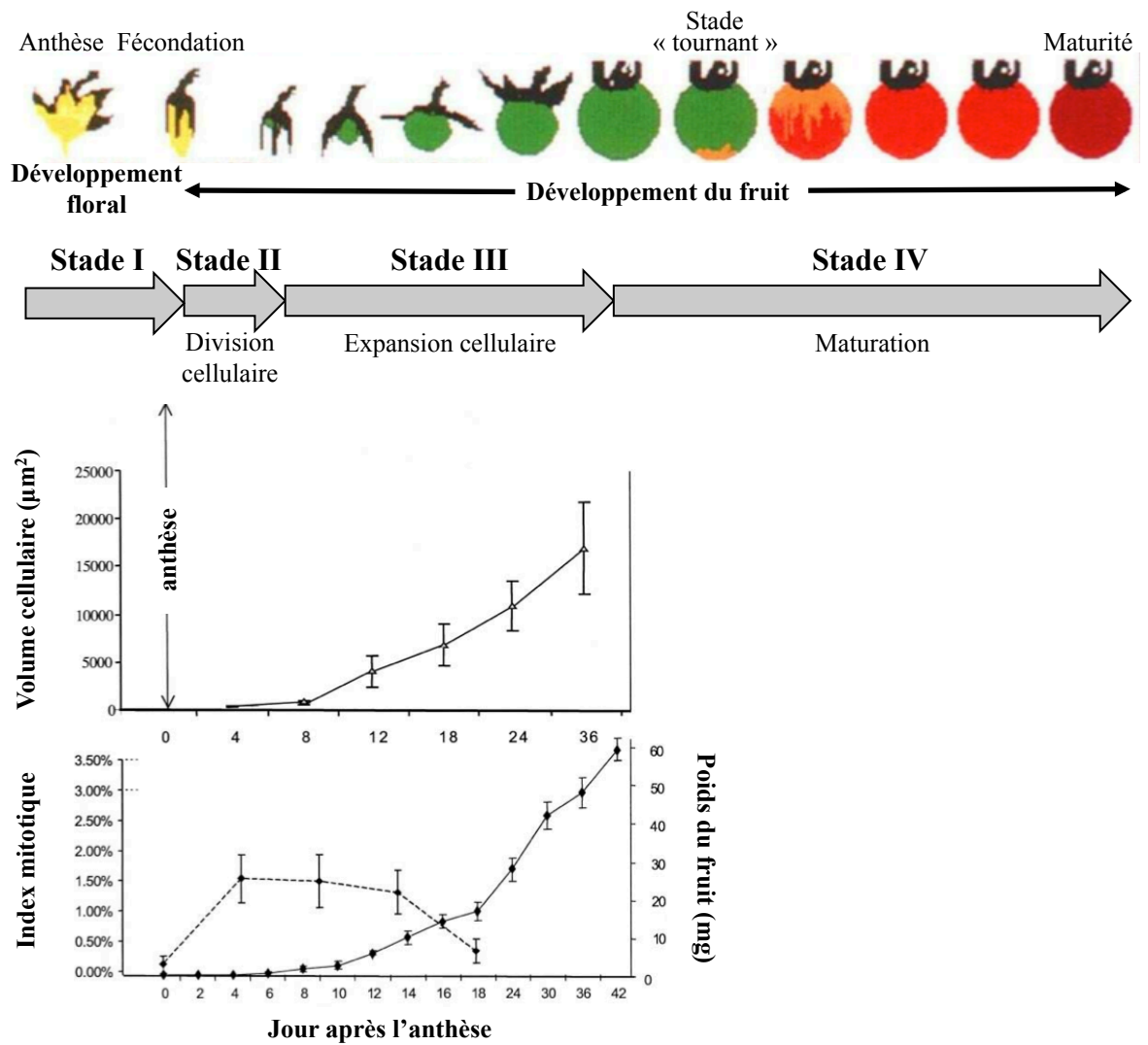


Andersen et al, 1995). Le pH de la baie est alors proche de 2,7 et la concentration en sucre est très faible (Terrier et al, 1996). Parallèlement, les ovules fécondés se développent en pépins qui atteindront leur poids et leur taille maximum à la fin de cette période;

- le stade II est caractérisé par un ralentissement de la croissance appelé plateau herbacé (Winkler et al, 1974) et dure une quinzaine de jours. La fin de cette phase, nommée véraison (pointée par une flèche noire; Figure 1.22) est caractérisée par le changement de couleur de la pellicule facilement identifiable chez les variétés rouges, plus difficilement chez les variétés blanches et le ramollissement palpable du péricarpe dû à une diminution de la concentration en substances pectiques (Silacci et Morrison, 1990) et à l'activité d'enzymes impliquées dans le remaniement des parois cellulaires (Donèche et Chardonnet, 1992; Barnavon et al, 2001). La véraison précède l'entrée dans la phase III d'environ 6 jours (Coombe, 1980; Ojeda et al, 1999). Avant véraison la baie est perçue comme un puits "métabolisant" alors qu'après véraison, durant la maturation, la baie est assimilée à un puits "stock" (Ollat et al, 2002);
- le stade III, dit de maturation, est le deuxième stade de croissance de la baie et découle exclusivement de l'augmentation du volume cellulaire par l'entrée de sucres et d'eau (Ojeda et al, 1999; Ollat et al, 2002). Les sucres sont l'*osmoticum* principal permettant le gonflement de la baie (Champagnol, 1984). En effet, les baies de vigne sont parmi les fruits les plus sucrés (Coombe, 1976) et la concentration en sucre peut dépasser une mole à maturité. Le glucose et fructose sont accumulés en quantités comparables et représentent 99% des sucres solubles (Winkler et al, 1974). De manière opposée, la concentration en acides organiques diminue et s'accompagne d'une augmentation du pH. A côté des effets de dilution dus au grandissement, des voies de dégradation ont été mises en évidence pour l'acide malique (Lakso et Kliever, 1975). La phase de maturation dure environ 40 jours mais sa fin est difficilement identifiable car elle varie d'un génotype à l'autre et dépend beaucoup de l'environnement. Sans qu'il s'agisse d'une règle absolue, le poids des baies double entre la véraison et la fin de la maturation. La pulpe représente alors de 64 à 90% du poids de la baie (Lavee et Nir, 1986).

### 1.3.2.3. Développement de la baie de tomate

Gillaspy et al (1993) et Tankley (2004) ont fait la synthèse de la plupart des résultats



**Figure 1.23:** Différents stades du développement de la baie de tomate (d'après Gillaspy et al, 1993 et Tankley, 2004).

obtenus sur le développement de la tomate, qui se décompose en quatre stades (Figure 1.23) :

- le stade I concerne la formation de l'ovaire à partir de l'initiation florale jusqu'à l'anthèse, et dure 2 à 3 semaines. A la fin de ce stade, l'identité, le nombre ainsi que la forme de chaque organe sont déterminés;

Les stades suivants correspondent au développement du fruit à partir de la fécondation de la fleur précédant l'anthèse d'environ 20 heures (Picken, 1984).

- le stade II se traduit par une activité mitotique intense au niveau de tous les tissus de l'ovaire, et dure environ deux semaines après la fécondation. Cheniclet et al (2005) ont également détecté une petite augmentation du volume cellulaire à ce stade du développement;
- le stade III correspond à la croissance du fruit exclusivement par extension cellulaire, qui commence dès que les divisions cellulaires cessent. Le volume cellulaire peut augmenter d'un facteur 22 000 pendant cette période (Cheniclet al, 2005). Cette augmentation est due à l'augmentation du niveau de ploïdie nucléaire des cellules du péricarpe, du placenta, de la columelle et du gel loculaire, phénomène appelé endoreduplication. La quantité d'ADN finale des cellules peut aller jusqu'à 256C (C étant la quantité d'ADN haploïde; Bergervo et al, 1996; Cheniclet et al, 2005). Mais, cette augmentation du volume cellulaire est également due à une entrée d'eau et d'acides organiques (principalement citrique et malique) dans la vacuole des cellules. Entre l'anthèse et la fin de ce stade, l'épaisseur du péricarpe peut augmenter d'un facteur 50 (Cheniclet et al, 2005). A la fin de ce stade, les graines sont entièrement formées;
- le stade IV correspond à l'arrêt de la croissance du fruit: la taille et la forme finale du fruit resteront inchangées jusqu'à maturation complète du fruit, tandis que d'autres changements s'opèrent tels que le changement de la couleur de la pellicule, de la texture de la chair, du rapport acide/sucre dans les cellules et un ramollissement général des parois cellulaires. La taille finale du fruit étant déterminée avant la phase IV, cette phase est parfois dissociée du développement de la tomate.

De grandes similitudes, mais également des différences existent donc entre le développement de la baie chez la vigne et chez la tomate (Tableau 1.6). En effet, le stade I chez la vigne, est caractérisé par une croissance de la baie due à de fortes divisions cellulaires puis à un élargissement des cellules et également par la formation des pépins. Ce stade correspond parfaitement aux stades II et III de la tomate durant lesquels les mêmes

**Tableau 1.6:** Comparaison entre les différents stades du développement de la baie de vigne et de tomate.

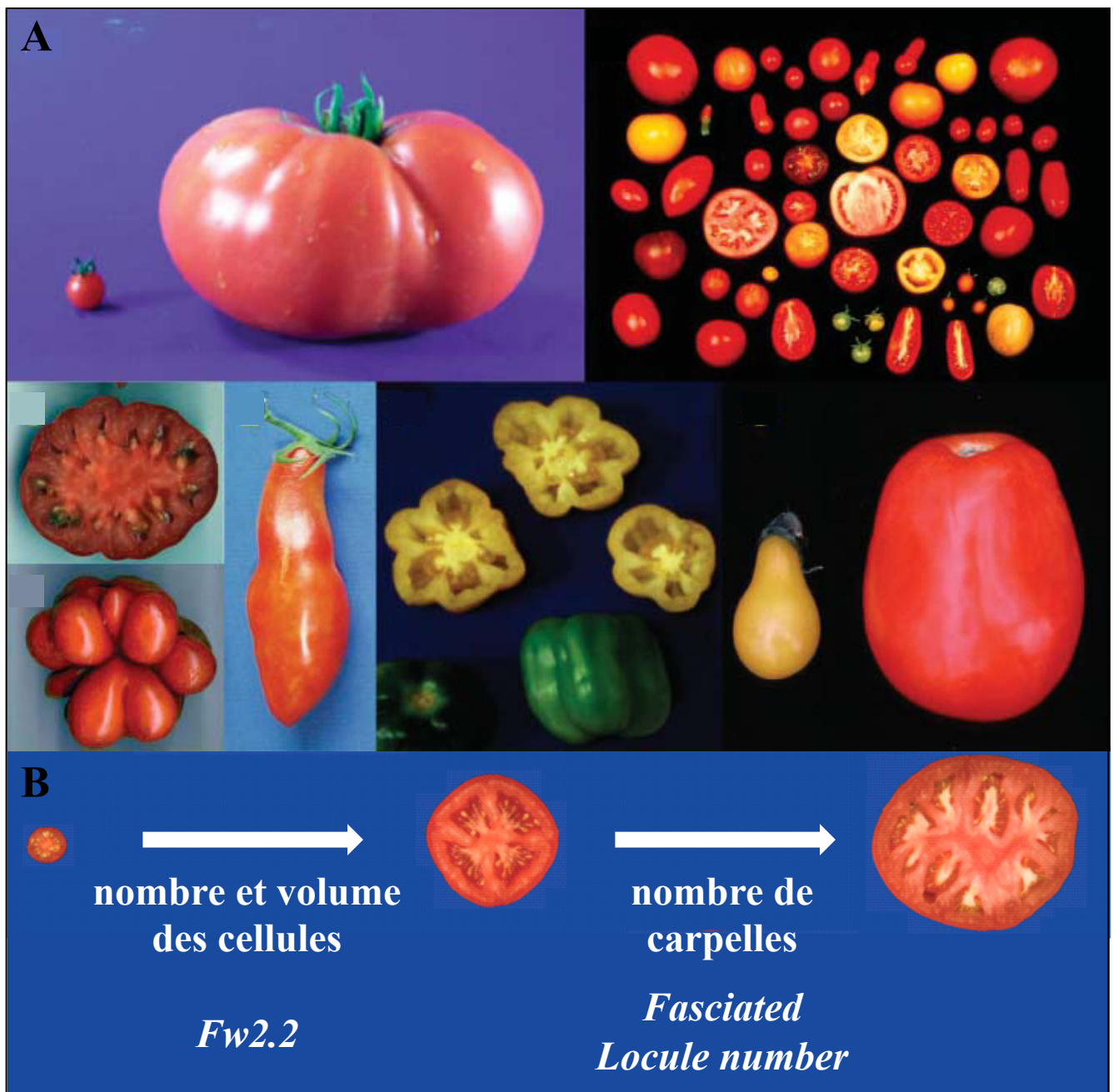
VIGNE		TOMATE	
Stade	Caractéristiques	Stade correspondant	Caractéristiques
I	Croissance par division cellulaire	II	Croissance par division cellulaire
	-----		
I	puis par expansion cellulaire (entrée d'eau et d'acide) Pépins entièrement formés	III	Croissance par expansion cellulaire (entrée d'eau et d'acide et endopolyplœdie) Pépins entièrement formés
II	Arrêt de la croissance - stade véraison		
III	Croissance par expansion cellulaire (entrée d'eau et de sucre) Maturation	IV	Arrêt de la croissance - stade « tournant » Changement des acides en sucres Maturation

phénomènes sont observés. Cependant, chez la tomate, l'augmentation cellulaire est due, en plus de l'entrée d'eau et d'acide dans la vacuole des cellules, à une augmentation de la ploïdie des cellules. Ce phénomène n'est pas observé dans les baies de vigne (Ojeda et al, 1999). Par ailleurs, à ce stade l'entrée d'eau se fait *via* le xylème chez la vigne (Ollat et al, 2002), alors qu'elle est d'origine phloémique chez la tomate (Gillaspy et al, 1993). Chez la tomate comme chez la vigne, la phase de maturation est précédée par un stade dans lequel on observe un arrêt de la croissance des fruits et un changement de métabolisme, notamment le renversement du ratio sucre/acide, ce stade appelé « véraison » chez la vigne et « tournant » chez la tomate est visualisable par un changement de couleur et de texture. Mais, à la différence de la tomate qui conserve le même volume jusqu'à la fin de la maturation, la baie de vigne reprend sa croissance après la véraison durant le stade III, *via* une expansion cellulaire due à une entrée d'eau et de sucres apportés par le phloème (Ollat et al, 2002). Cette reprise de la croissance est à l'origine de la forme en double sigmoïde de sa courbe de croissance observée chez la vigne, mais pas chez la tomate. Enfin, on peut également noter d'autres facteurs divergents entre le développement de la baie de vigne et de tomate, comme l'origine du péricarpe qui dérive de la couche méristématique L3 chez la tomate (Gillaspy et al, 1993) et L2 chez la vigne (Thompson and Olmo, 1963).

La taille finale de la baie chez la vigne et la tomate est donc dépendante *in fine* du nombre de cellules présentes dans cet organe ainsi que de leurs tailles, même si ces deux caractères n'ont pas le même facteur multiplicatif au cours du développement. En effet, Coombe (1976) a montré que chez la vigne, le volume des cellules de la pulpe pouvait augmenter d'un facteur 300 ou plus, alors que le nombre de cellules doublait seulement au cours du développement de la baie. De même chez la tomate, la taille des cellules augmente considérablement pendant le développement du fruit. La croissance de la baie dans ces deux espèces semble donc surtout liée au grandissement cellulaire. Cependant, des études génétiques sont nécessaires pour préciser l'importance respective de ces deux facteurs dans la variation de la taille du fruit.

### **1.3.3. Etude des déterminants de la variabilité de la taille des fruits charnus**

La taille des baies chez la vigne et plus généralement chez les fruits charnus, dépend de la variété (génotype), mais également de l'environnement dans lequel la plante s'est développée (Carbonneau et al, 2007; Carmona et al, 2008). Ces facteurs pouvant être le climat,



**Figure 1.24:** Variation de la taille du fruit chez la tomate. **(A)** Illustration de la gamme de variation de la taille, de la couleur et de la forme de la baie chez la tomate (Tankley, 2004). **(B)** Evolution de la taille de la tomate et gènes clonés impliqués (Cong, 2008).

le sol (température, lumière, CO<sub>2</sub>, pluie, humidité de l'air, disponibilité des sources de carbone, composants du sol), les pratiques viticoles par leur influence sur la vigueur et la charge en fruit (apport de nutriments, irrigation, choix du porte greffe, entretien du sol, mode de taille) et également le nombre de fleurs fécondées (le nombre de baies sur la grappe), la position de la baie dans la grappe... Au cours de ma thèse, je ne me suis intéressée qu'à la composante génétique de la variation de la taille des baies, dans un dispositif expérimental qui ne permet pas de prendre en compte la composante environnementale.

Je présente ici l'état des connaissances chez la tomate et la vigne concernant les déterminants physiologiques et moléculaires de la variation de la taille du fruit chez ces deux espèces.

### **1.3.3.1. Variation de la taille de la baie chez la tomate**

#### **1.3.3.1.1. Déterminants physiologiques**

La sélection des plantes représente une forme accélérée de l'évolution conduisant à des changements exagérés des tissus et organes d'intérêt pour l'Homme (fruit, graine, racine, tubercule...). Un des cas les plus extrêmes et le mieux étudié est la variation de la taille de la tomate (Tansley, 2004). En effet, les tomates sauvages hypothétiquement ancêtres des variétés actuelles, produisent des petits fruits ronds pesant seulement quelques grammes alors que les tomates des variétés modernes peuvent peser jusqu'à 1 kg (Figure 1.24-A). De nombreuses études portant sur l'analyse de la forme, du nombre de carpelles ou de la taille et du volume cellulaire du fruit, ont été menées afin d'identifier les mécanismes sous-jacents à la variation de la taille du fruit chez la tomate.

Ces études ont été réalisées sur des descendance de RIL contenant des introgressions de tomates sauvages (3472 plantes, Frary et al, 2000 ; 3123 plantes, Cong et al, 2008). Enfin, d'autres études ont utilisé des lignées recombinantes pour la détection de QTL de taille du fruit, lignées dérivées du croisement entre deux variétés contrastées pour la taille du fruit (2 ou 4 lignées, Bertin et al, 2003 et 2009), ou bien une collection de tomates sauvages et cultivées contrastées pour la taille de la baie (20 génotypes, Cheniclet et al, 2005; 360 génotypes, Ranc, 2010).

L'ensemble de ces études a permis de mettre en évidence que chez la tomate trois paramètres principaux contribuent à la variation de la taille du fruit entre génotypes:

- l'élargissement cellulaire (Cheniclet et al, 2005): en effet, le volume cellulaire peut varier d'un facteur 35 entre les variétés. Cheniclet et al (2005) ont montré que le

volume cellulaire était corrélé positivement avec le niveau de ploïdie de la cellule. Cependant, les cellules du fruit de certaines variétés ne présentent pas de variation de la ploïdie mais présentent bien une variation de 30% de la taille des cellules du péricarpe (Bertin et al, 2003). La variation du volume cellulaire peut alors s'expliquer par un deuxième mécanisme qui correspond à une augmentation du volume de la cellule par l'entrée de solutés et d'eau;

- le nombre de cellules du péricarpe (Bohner et Bangerth, 1988; Frary et al, 2000; Bertin et al, 2003 et 2009). Il existe deux mécanismes de divisions cellulaires conduisant à des différences de taille du fruit: les divisions périclinales et des divisions non-orientées. Frary et al (2000) ont montré que 30% de la variation de la taille des fruits était due à une différence globale du nombre de cellules avant et après l'anthèse liée à des divisions non-orientées. Cependant, il a également été montré que des variétés présentant un nombre de couches cellulaires du péricarpe similaire à l'anthèse pouvaient par la suite gagner entre 5 à 17 couches cellulaires supplémentaires par division péricinale des cellules, créant ainsi des différences au niveau de l'épaisseur du péricarpe (Cheniclet et al, 2005; Bertin et al, 2003);
- le nombre de carpelles (Cong et al, 2008; Figure 1.24-A): ce nombre varie de deux (chez les tomates sauvages et variétés à petits et moyens fruits) à plus de vingt, chez les variétés à très gros fruits (exemple: cultivar Gros de Gros pesant jusqu'à 500 g; Cheniclet et al, 2005) et explique jusqu'à 50% de la variation du poids final entre variétés (Cong et al, 2008).

Le poids final du fruit étant dépendant du nombre de carpelles, la relation entre la taille des cellules ou le nombre de cellules et le poids final du fruit n'est juste que si l'on compare des variétés ayant un nombre de carpelles similaires. Tanksley (2004) et Cong et al (2008) suggèrent que la progression de la taille de la tomate, de la domestication à aujourd'hui, serait d'abord due à une augmentation du nombre des cellules du péricarpe puis à une augmentation du nombre de carpelles (Figure 1.24-B). Enfin, chez la tomate mais également chez d'autres fruits (Coombe, 1976 ; Tanksley, 2004), l'augmentation de la taille du fruit chez les variétés cultivées est accompagnée d'une variation phénoménale de la forme (Figure 1.24-A), le nombre et le type de divisions cellulaires (périclines ou anticlines) contribuant au contrôle de la forme et de la taille finale du fruit.



### 1.3.3.1.2. Déterminants génétiques et moléculaires

De très nombreuses études génétiques sur la variation de la taille du fruit ont été menées chez la tomate, montrant que ce caractère est héritable et quantitatif (Tanksley et al, 2004; Gupsa et al, 2009; Lozano et al, 2009 pour revues). Environ 30 QTL seraient impliqués dans la variation du poids et/ou de la forme de la tomate (Grandillo et al, 1999; Causse et al, 2004) dont moins de 10 ayant des effets majeurs sur le poids et/ou la taille du fruit (Grandillo et al, 1999) et sont listés ci-après :

- quatre QTL majeurs sont responsables d'une part importante de la variation du poids du fruit, indépendamment du nombre de carpelles: *Fw1.1*, *Fw2.2*, *Fw3.1* et *Fw4.1*. L'un d'entre eux, *Fw2.2*, qui explique 30% de la variation du poids final du fruit, a été identifié (Frery et al, 2000). *Fw2.2* code une protéine similaire à un oncogène humain. Ce gène serait un régulateur négatif de la division cellulaire pendant les phases précoces du développement du fruit (Frery et al, 2000; Cong et al, 2002) et régulerait le nombre global de cellules du péricarpe mais pas le nombre de couches cellulaires (Cong et al 2003; Liu et al, 2002). Ce résultat suggère donc un contrôle génétique différent pour ces deux modes de divisions. Cependant, aucun gène impliqué dans le second mode de division cellulaire n'a encore été identifié;
- deux QTL responsables de l'augmentation du nombre de carpelles dans l'ovaire ont été identifiés: *Fasciated* et *Locule-number*. *Fasciated* expliquerait jusqu'à 50% de la variation du poids final du fruit (Cong et al, 2008). *Locule-number* a un effet beaucoup moins fort que *Fasciated* (3 à 4 loges pour *Locule-number* et >15 pour *Fasciated*). Cependant, l'interaction des deux allèles favorables à ces loci aboutit à des fruits avec un nombre encore plus élevé de carpelles et pesant plus de 500 g (Tanksley, 2004). Les gènes sous-jacents à ces deux QTL ont été récemment identifiés. Le gène *Fasciated* code un facteur de transcription de type YABBY dont la sous-expression dans les *primordia* carpellaires (en raison d'une insertion dans son premier exon) conduit à la formation d'un grand nombre de carpelles (Cong et al, 2008). La variation sous-jacente au QTL *Locule-number* correspond à la présence de 2 SNP dans une région en amont de l'orthologue de séquence du gène *Wuschel* (facteur de transcription de type homeobox impliqué dans l'identité du méristème florale chez *Arabidopsis*) mais aucune modification de l'expression de ce gène n'a été identifiée entre des cultivars phénotypiquement différents pour le nombre de loges, les mécanismes moléculaires impliqués restent à élucider (Ranc, 2010). Les QTL



*Fasciated* et *Locule-number* sont également associés à une variation de la forme du fruit car l'augmentation du nombre de carpelles tend à aplatir les fruits et à leur donner une apparence côtelée;

- trois QTL nommés *Ovate*, *Sun* et *Fs8.1* sont responsables d'une part importante de la variation du rapport entre la longueur du fruit et sa largeur (60% de la variation expliquée par *Sun* et 40% par *Ovate* dans deux populations différentes). Bien que les loci *Ovate* et *Sun* soient tous les deux impliqués dans le passage de la forme ronde à allongée du fruit, il existe des différences morphologiques de l'effet de ces QTL. Le QTL *Ovate* crée une élongation asymétrique avec une élongation du haut du fruit plus importante que celle du bas (forme pyriforme) alors que l'effet de *Sun* est uniforme (forme ovoïde). *Fs8.1*, quant à lui, induit une forme « carré » du fruit. Seuls les gènes sous-jacents aux QTL *Ovate* et *Sun* ont été identifiés. *Ovate* code une nouvelle classe de protéines régulatrices présentes dans le noyau (Liu et al, 2002). La mutation responsable du caractère allongé du fruit est provoquée par un codon stop dans le deuxième exon de ce gène. *Sun*, quant à lui, code une protéine liant le calcium, qui lorsqu'elle est abondante serait à l'origine de l'allongement du fruit (Xia et al, 2008). La fonction exacte de ces gènes dans le développement du fruit reste cependant à élucider.

Enfin, des études ont suggéré que l'interaction entre deux ou plusieurs de ces loci pourrait expliquer les phénotypes les plus extrêmes en terme de forme et de taille du fruit. Par exemple, la variété *Giantheirloom*, qui produit de très gros fruits, pouvant peser jusqu'à 1 000 g, doit ce phénotype à l'action combinée d'au moins trois gènes présentés précédemment : *Fw2.2*, *Locule-number* et *Fasciated* (Lippman et Tanksley, 2001). De même, la variété *Yellow Stuffer* ressemblant à un poivron doit sa forme principalement à une interaction entre *Fw2.2*, *Fs8.1* et *Locule-number* (Van der Knapp et Tanksley, 2003).

Il faut aussi noter qu'indépendamment des études QTL, plusieurs protéines impliquées dans la variation de la taille du fruit chez la tomate ont été identifiées. Une 3-hydroxy-3-methylglutaryl-coenzyme A réductase (HMGR), protéine impliquée dans la biosynthèse d'isoprène, précurseur de plusieurs phytohormones, présente une activité et un niveau de transcrits élevé pendant les stades précoces du développement de la tomate (Nariat et Gruissem, 1989). De plus, une inhibition de l'activité *in vivo* de cette protéine (par ajout de mévinoline), pendant les stades précoces du développement, perturbe le développement du fruit et aboutit à une réduction de leur taille par rapport à des fruits non traités, alors qu'une



**Figure 1.25:** Illustration de la gamme de variation de la taille et de la forme de la baie chez la vigne cultivée.

inhibition à des stades plus tardifs n'a pas d'effet (Nariat et Gruissem, 1989). Le rôle de cette protéine sur le contrôle de la taille ou du nombre de cellules du fruit chez la tomate a également été suggéré par Jelesko et al (1999). Deux autres protéines impliquées dans la variation du volume cellulaire *via* l'endoréduplication ont été identifiées chez la tomate (Chevalier et al, 2011 pour revue). La protéine kinase WEE1 (Gonzalez et al, 2007) inhibant l'activité des kinases cycline-dépendantes (CDK) et la protéine CCS52A, un activateur de voie de dégradation des cyclines mitotiques (Mathieu-Rivet et al, 2010). Ces deux protéines participent au contrôle de la taille des cellules par une augmentation de la ploïdie due à une perturbation du cycle cellulaire pendant le développement précoce de la tomate. Cependant, aucun gène impliqué dans la variation de la taille des cellules, *via* l'entrée de soluté (acide/sucre), n'a encore été identifié à ce jour, bien que de nombreux QTL impliqués dans la variation du taux d'acidité et de sucre de la baie ont été cartographiés dont certains co-localisent avec des QTL pour la taille du fruit (Tankley et al, 1996; Fulton et al, 2002; Causse et al, 2004).

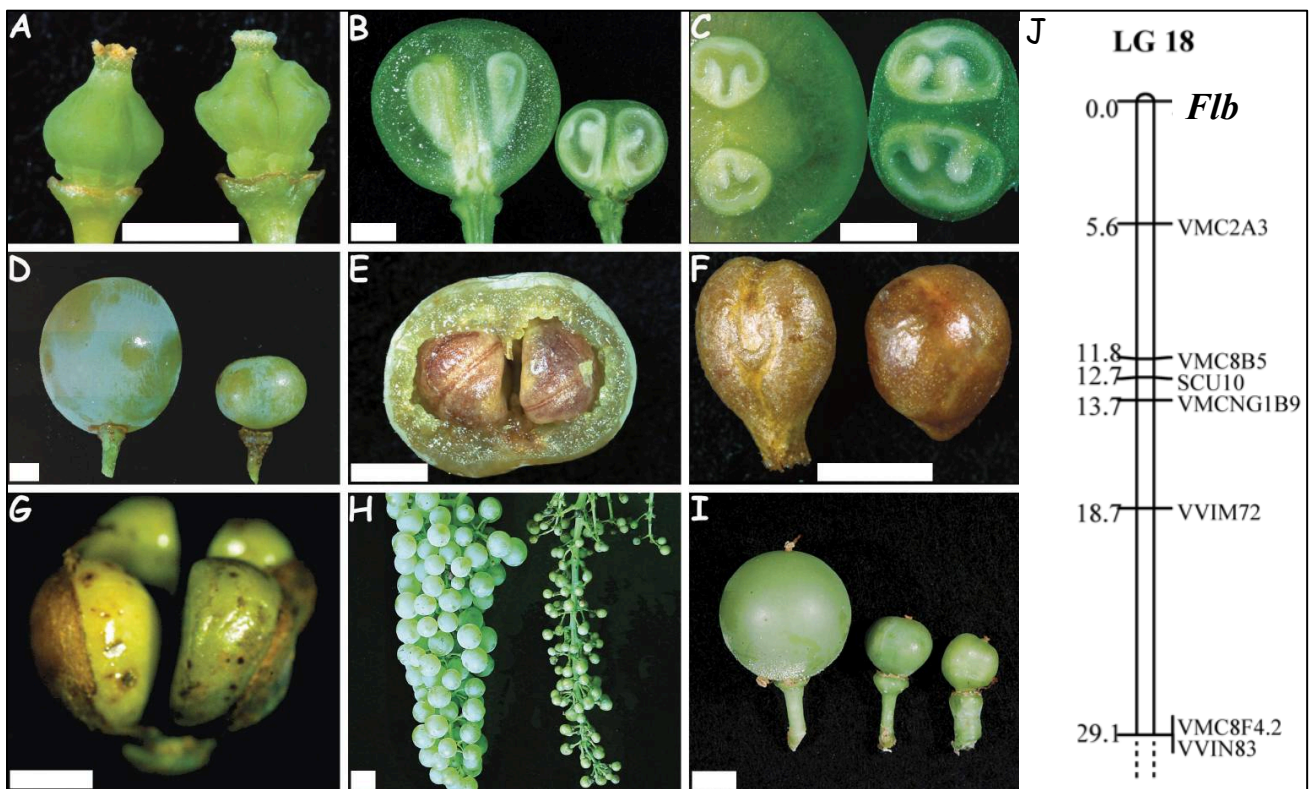
Le déterminisme génétique et moléculaire de la variation du fruit chez la tomate reste donc encore à compléter. De plus, l'ensemble de ces résultats montre la complexité des mécanismes impliqués dans la variation de la taille de la tomate et certainement des fruits charnus en général.

### **1.3.3.2. Variation de la taille de la baie chez la vigne**

Chez la vigne, le poids mais également la forme de la baie présentent une forte variabilité entre les variétés (Figure 1.25; Galet et al, 2000), mais également entre clones d'une même variété (Fernandez et al, 2006a et 2006b). En effet, Boursiquot et al, (1995), ont montré que la taille des baies varie de 0,5 g à 11 g dans le compartiment cultivé (données d'observations de 5311 génotypes de *V. vinifera* ssp *sativa* de la collection du Domaine de Vassal). Cependant, malgré l'importance de ce caractère pour l'amélioration des raisins de table et la qualité de la vendange (Champagnol, 1998), les déterminants de la variabilité de ce caractère sont bien moins connus que chez la tomate.

#### **1.3.3.2.1. Déterminants physiologiques**

Chez la vigne cultivée, deux études se sont intéressées à décomposer la variabilité de la taille des baies en fonction du nombre et de la taille des cellules de la pulpe. La première porte sur l'analyse de six génotypes de *V. vinifera* cultivées présentant une variabilité



**Figure 1.26:** Morphologie du mutant *Fleshless berry*. Morphologie des ovaires (A) , des baies (B, C, D, E, H, I),et des pépins (F, G) de l’Ugni Blanc mutant *Fleshless berry* (à droite) comparés à ceux de l’Ugni Blanc non-muté à gauche. Cartographie génétique (J) de la mutation dans une descendance Chardonnant x Ugni Blanc mutant (Fernandez et al, 2006b et 2006c).

importante de la taille des baies: deux clones de la variété Grenache, deux clones de la variété Mourvèdre et deux clones de la variété Ugni Blanc (un mutant naturel nommé *Fleshless berry* (*Flb*) et l'Ugni Blanc non muté; Fernandez et al, 2006a). Dans les deux premières variétés, une différence de taille des cellules de la baie est à l'origine de la variation de la taille du fruit observée entre les deux clones de chaque variété (estimation indirecte du volume cellulaire et du nombre de cellules à partir de la quantité d'ADN de la pulpe d'après Ojeda et al, 1999) alors que pour la variété Ugni Blanc le nombre et la taille des cellules de la baie sont à l'origine de la différence de taille des baies entre les deux clones. La seconde, est une étude histologique plus approfondie du mutant *Fleshless berry* et de l'Ugni Blanc non muté (Fernandez et al, 2006b), les coupes histologiques réalisées confirment que le volume des cellules mais également leur nombre sont à l'origine de la variation de la taille du fruit entre les deux clones et que cette différence apparaît après l'anthèse. Ce mutant est un outil précieux pour l'étude de la variation de la taille du fruit car il présente une réduction par 20 du volume du péricarpe par rapport à l'Ugni Blanc non muté (Figure 1.26 ; Fernandez et al, 2006b). Cette différence est liée à une réduction de l'expansion cellulaire et à une altération des divisions des cellules du mésocarpe interne qui sont à l'origine de la pulpe de la baie. Le septum, également à l'origine de la pulpe de la baie chez la vigne, n'est en revanche pas touché par la mutation et participe à la pseudo-pulpe du mutant, suggérant un contrôle génétique différent pour le développement de ces deux tissus. La différence de la taille des baies entre les deux clones apparaît après l'anthèse puisqu'à l'anthèse les deux clones présentent la même organisation des tissus de l'ovaire, avec un nombre et un volume des cellules identiques avec cependant, une paroi irrégulière avec une ride chez les ovaires du mutant (Figure 1.26-A). La mutation ne semble pas avoir d'effet sur la fertilité ou sur le nombre et la taille des pépins. Des caractéristiques similaires ont été retrouvées chez les mutants *Ful-1* d'*Arabidopsis* (Gu et al, 1998) et dans les lignées de tomate possédant l'allèle *Fw2.2* (Liu et al, 2003). Ce dernier résultat indique que l'effet de la mutation sur le phénotype du fruit correspond uniquement à une altération des tissus maternels du fruit et suggère que la formation d'embryon zygotique fonctionnel et le développement du fruit sont deux processus physiologiquement indépendants. Cependant, le mutant présente des pépins plus arrondis que le génotype non muté (Fernandez et al, 2006b; Figure 1.26-F). De par leur taille réduite et leurs pépins ronds, les baies du mutant *Flb* ressemblent beaucoup à celles des génotypes *V. vinifera* ssp *silvestris* (Figure 1.6). En effet, les pépins de forme arrondie sont une des caractéristiques des vignes sauvages (This et al, 2006; Terral et al, 2010). Enfin, d'autres différences existent entre les deux clones comme une diminution importante de la





concentration en acides organiques (malate, tartrate et citrate) et en sucres (glucose, sucrose et fructose) pendant le développement de la baie mais ces changements ne sont pas suffisants pour permettre une modification aussi significative dans la taille de la baie (Fernandez et al, 2006b).

D'autres études, en relation avec la variation de la taille de la baie, ont été réalisées dans des descendance pleins-frères ségrégant pour la taille de la baie mais également pour l'apyrénie (Doligez et al, 2002; Fanizza et al, 2005; Cabezas et al, 2006; Mejia et al, 2007; Constantini et al, 2008), mais le phénotypage des baies reposait exclusivement sur le poids et/ou le diamètre des baies, le poids et le nombre de pépins. Ces études montrent qu'il existe une corrélation entre le nombre de pépins et la taille finale des baies, cette corrélation étant certainement due à la production de gibbérellines (GA) par les pépins. En effet, Scienza et al (1978) ont observé un pic de GA durant la première période de croissance des baies et une corrélation entre la concentration de GA et le nombre de pépins. Cependant, cette corrélation entre le nombre de pépins et la taille des baies, en relation avec la production de GA, serait spécifique des génotypes apyrènes ou partiellement apyrènes. L'application de GA est une pratique très répandue pour stimuler la croissance de la baie des variétés apyrènes. En effet, une augmentation de la taille des baies est bien observée chez le raisin de Corinthe et la Sultanine, deux cultivars (*cv.*) apyrènes, en réaction à des traitements exogènes de GA au moment de la floraison ou peu de temps après (Coombe, 1973). En revanche, la taille des baies de cultivars à pépins n'est pas affectée par les traitements aux GA (Weaver et Pool, 1965). De plus, le nombre de pépins n'est pas la seule explication des variations de taille du fruit observées entre différentes variétés. En effet, malgré une différence de la taille moyenne des baies à maturité du *cv.* Merlot et du *cv.* Cabernet Sauvignon, respectivement 2 et 1,6 g, la proportion de baies à un, deux et trois pépins, est identique dans les deux variétés (Ollat et al, 2002). Il existe donc d'autres mécanismes expliquant la taille des baies entre génotypes.

Enfin, chez la vigne sauvage, le poids de la baie a été très peu étudié, seules deux études ont comparé le poids de la baie entre génotypes (Levadoux, 1956; Scienza et al, 1985) et aucune étude prenant en compte le nombre ou la taille des cellules des baies n'a été réalisée.

A l'heure actuelle, aucune description détaillée des différences morphologiques impliquées dans la variation de la taille de la baie n'a été réalisée sur un échantillon représentatif de la vigne cultivée ou sauvage afin de confirmer les résultats obtenus précédemment, et aucune étude ne s'est intéressée au rôle des carpelles dans la variation de la

**Tableau 1.7:** Ensemble des études QTL pour le caractère poids de la baie chez la vigne cultivée.

Population	Nombre de marqueurs	Nombre d'individus	Années de phénotypage	Nombre de QTL poids de baie <sup>§</sup>	co-localisation QTL poids baie/ nombre de pépins	Référence
F1: (Dattier de Beyrouth x Pirovano 75 <sup>§</sup> ) x (Alphonse Lavallée x Sultanine <sup>§</sup> )	301	139	3	1 (chromosome: 18***)	oui (chromosome 18)	Doligez et al, 2002
F1: Italia x Big Perlon <sup>§</sup>	313	184	3	5 (chromosomes: 4, 5, 13, 16, 20)*	aucun	Fanizza et al, 2005
F1: Dominga x Autumm Seedless <sup>§</sup>	595	118	3	2 (chromosomes: 15, 18)**	oui (chromosome 18)	Cabezas et al, 2006
F1: Ruby Seedless <sup>§</sup> x Sultanine <sup>§</sup>	154	144	2	2 (chromosomes: 15, 18)**	oui (chromosome 18)	Mejia et al, 2007
F1: Italia x Big Perlon <sup>§</sup>	341	163	3	3 (chromosomes 1, 12, 18)***	oui (chromosome 18)	Costantini et al, 2008
F1: Ruby Seedless <sup>§</sup> x Sultanine <sup>§</sup>	63	139	3	3 (chromosomes 2, 8, 18)***	oui (chromosome 18)	Mejia et al, 2011 et résultats non publiés

<sup>§</sup> variété apyrène

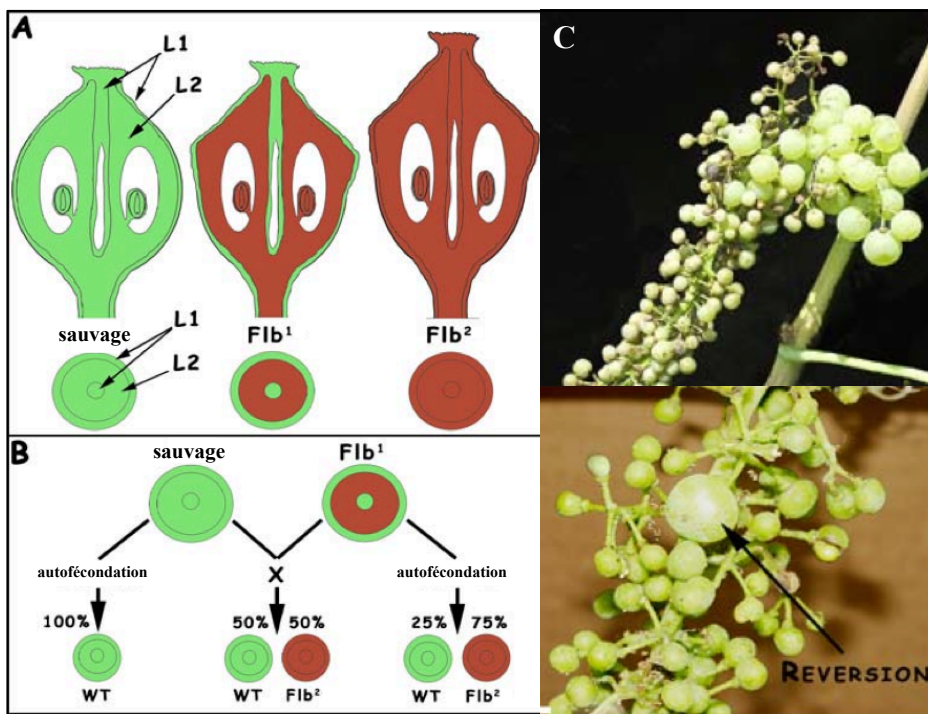
<sup>§</sup> détection du QTL sur \*1 année \*\*2 années \*\*\*3 années

taille de la baie. Enfin, aucune étude de QTL sur la variation ou l'accumulation du taux de sucre ou de l'acidité n'a encore été réalisée, cette étude permettrait de préciser l'implication des différents solutés dans le mécanisme de grandissement cellulaire et donc dans le grossissement du fruit.

#### 1.3.3.2.2. Déterminants génétiques et moléculaires

On ne dispose pour l'instant que de quelques études de détection de QTL pour le poids de la baie dans des croisements avec des variétés apyrènes (Tableau 1.7; Doligez et al, 2002; Fanizza et al, 2005; Cabezas et al, 2006; Mejia et al, 2007; Costantini et al, 2008 ; Mejia et al, 2011). L'ensemble de ces études a permis d'identifier 12 QTL pour le poids de la baie dont seulement un est commun à cinq des six études (QTL non identifié par Fanizza et al, 2005). Ce QTL, se trouve sur le groupe de liaison 18, explique jusqu'à 40% de la variation de la taille de la baie et co-localise avec le QTL majeur de l'apyrénie (Mejia et al, 2007; Costantini et al, 2008; Cabezas et al, 2006). Cette co-localisation pourrait s'expliquer soit par l'effet pleiotropique du gène sous-jacent sur les deux caractères, soit par l'existence de deux QTL très proches dans la région. La première hypothèse semble la plus vraisemblable, dans la mesure où une relation positive entre le nombre de pépins et la taille de la baie a été observée dans ces mêmes études (Doligez et al, 2002 ; Cabezas et al, 2006; Mejia et al, 2007; Costantini et al, 2008 ; Mejia et al, 2011). Les autres QTL identifiés ont en général un effet mineur et indépendant du nombre de pépins, montrant qu'il existe des facteurs indépendants pour les deux caractères. Cependant, ces QTL sont différents entre les études et montrant une instabilité en fonction des fonds génétiques et des années de phénotypage (Tableau 1.7). L'ensemble de ces résultats illustre la difficulté à identifier les déterminants génétiques de la variabilité de la taille de la baie chez la vigne. Principalement en raison de la lourdeur des dispositifs expérimentaux qui seraient nécessaires au contrôle des effets environnementaux, mais également au choix des descendances qui ségrégent en majorité aussi pour le caractère apyrène en forte interaction avec le caractère taille de la baie dans ces études.

Un autre type d'approche basé sur l'étude du mutant *Fleshless berry* a été réalisé. Afin d'analyser l'héritabilité de la mutation *Flb*, plusieurs populations ont été produites: des autofécondations du mutant et du génotype non muté et des croisements réciproques Ugni B mutant x Chardonnay et Chardonnay x Ugni B mutant (Fernandez et al, 2006c). L'analyse de ces populations a révélé que la mutation *Flb* était hétérozygote, dominante et certainement



**Figure 1.27:** Organisation chimérique du mutant *Fleshless berry*. (**A et B**) Schéma des couches L1 et L2 mutées (rouge) ou non (verte) et du phénotype des baies correspondantes. (**C**) Photo de réversion du phénotype *Flb* chez un individu de la descendance Chardonnay x Ugni Blanc mutant (Fernandez et al, 2006c).

monogénique. Mais le résultat le plus surprenant a été le phénotype de certains descendants. En effet, les ovaires des certains descendants portant un allèle *Flb* sont 20 à 50% plus gros (Figure 1.27-A), ne se développent pas en fruit et sont stériles. Ce phénotype a été nommé  $Flb^2$  par les auteurs, pour faire la comparaison avec le phénotype du mutant original  $Flb^1$ . Les auteurs montrent que le phénotype  $Flb^1$  s'expliquerait par un statut chimérique de la mutation *Flb* dans le mutant original qui serait présente seulement dans la couche L2 de cellules du méristème apical (Figure 1.27-A), mais pas dans les cellules du stigmate et des tissus internes du style de l'ovaire, qui dérivent de la couche cellulaire L1. Les descendants d'autofécondation du mutant ou du croisement du l'Ugni Blanc non muté et du mutant présentent le phénotype  $Flb^2$  respectivement dans 75% et 50% des cas (Figure 1.27-B), du à l'hétérozygotie de l'allèle muté dans le parent d'origine et de la dominance de l'allèle muté. La présence de la mutation dans les tissus L1 (stigmates et tissus internes du style de l'ovaire) entraîne l'arrêt complet du développement du fruit, expliquant le phénotype  $Flb^2$  (Fernandez et al, 2006c). Une instabilité de la mutation *Flb* a également été observée. En effet, des réversions se manifestant par l'apparition d'une ou plusieurs baies de taille normale sur une grappe d'une plante *Flb* avec une fréquence de  $10^{-4}$  à  $10^{-3}$  (Figure 1.27-C). Deux hypothèses peuvent expliquer ces résultats: l'effet d'un élément transposable (comme pour la mutation du gène *MybA1* responsable de la variation de couleur de la pellicule ou du gène *VvTFL1A* impliqué dans la ramification de la grappe; Yakushiji et al, 2006; Fernandez et al, 2010), soit une instabilité de la couche L2 mutée qui pourrait être envahie par une couche L1 non mutée (Fernandez et al 2006c). Afin de cartographier la mutation, une population issue du croisement entre le Chardonnay et l'Ugni Blanc mutant a été réalisée (Fernandez et al, 2006c). L'analyse de la descendance F1 (71 individus) a permis de cartographier la mutation sur l'extrémité supérieure du groupe de liaison 18 après deux années successives de phénotypage (Figure 1.26-C) à 5,6 cM du marqueur le plus proche, VMC2A3. Cependant, la position du gène muté demanderait à être confortée, car il se trouve à l'extrémité d'un groupe de liaison.

Afin d'identifier des gènes potentiellement impliqués dans la formation du fruit et plus précisément de la pulpe, l'expression différentielle des gènes entre les deux clones d'Ugni Blanc a été analysée à l'aide d'une puce 3K et a permis d'identifier 53 et 50 gènes comme étant respectivement sur-exprimés ou sous-exprimés chez le mutant, une semaine après l'anthèse (Fernandez et al, 2007). En parallèle une hybridation soustractive a été réalisée à partir d'ARN extraits de 0 à 15 jours après l'anthèse entre le mutant et le sauvage et a permis l'identification de transcrits différentiellement exprimés pendant les premiers stades du

développement du péricarpe. Ces transcrits correspondent à des gènes impliqués dans le murissement, la réponse aux stress mais également à de nombreux facteurs de transcription comme *VvPISTILLATA*, *VvYABBY2*, *VvHB13* dont certains font l'objet d'une étude fonctionnelle par génétique inverse (L. Torregrosa communication personnelle).

L'identification des déterminants génétiques et moléculaires de la variation de la taille de la baie chez la vigne, en est donc à ses balbutiements. A l'heure actuelle aucun gène capable d'expliquer ne serait-ce que partiellement la variabilité observée au niveau du poids des baies n'a encore été identifié. Cependant, la disponibilité de la séquence du génome (Jaillon et al, 2007) et une meilleure exploitation du matériel génétique disponible, dont le mutant *Fleshless berry*, devrait permettre de progresser plus rapidement que par le passé.

## 1.4. Contexte et objectifs de la thèse

Le caractère taille de la baie est un facteur essentiel à la qualité des raisins de table mais aussi des raisins de cuve, car en agissant sur le rapport pulpe/pellicule, il conditionne la composition de la vendange en métabolites secondaires (Champagnol, 1998). C'est d'ailleurs, un des caractères de diversification dans le compartiment cultivé de la vigne (Boursiquot et al, 1995). Néanmoins, les mécanismes sous-jacents comme le nombre de cellules, la taille des cellules ou le nombre de carpelles n'ont pas été étudiés ou bien seulement chez quelques génotypes (Fernandez et al, 2006a et 2006b). Aucun gène impliqué dans la variation de la taille de la baie chez la vigne n'a encore été identifié à ce jour. Les travaux les plus avancés sur l'identification de déterminants génétiques de la taille des fruits charnus concernent la tomate (Tankley, 2004), mais il n'est pas complètement évident que ces connaissances puissent être directement transférables chez la vigne sachant que le développement de ces deux fruits présente des différences notables (endoréduplication, participation de la columelle et du placenta,...). Les connaissances actuelles des mécanismes génétiques impliqués dans la variation de la taille de la baie chez la vigne sont largement insuffisantes pour manipuler efficacement ce caractère dans des programmes de sélection.

Les objectifs de mon travail de thèse étaient de contribuer à l'identification des déterminants de la variation de la taille de la baie chez *V. vinifera*, de trois manières différentes :

- la première stratégie a consisté à étudier un grand nombre de paramètres des ovaires et des baies de vigne, à différents stades de développement dans une large collection de variétés cultivées, dans l'objectif d'identifier les caractères les plus importants et de tester la faisabilité de leur phénotypage à grande échelle;
- la deuxième stratégie a consisté à rechercher des associations entre la variation de ces caractères et des variations de séquences dans des gènes candidats puis de renforcer les résultats obtenus par l'analyse de l'expression de ces gènes au cours du développement de la fleur et de la baie entre des génotypes contrastés pour la taille de la baie;
- enfin, la troisième approche a consisté à analyser la diversité nucléotidique le long de la région portant le locus *Fleshless berry* entre des génotypes sauvages et cultivés, afin de détecter d'éventuelles traces de domestication chez les vignes cultivées en rapport avec la variation de la taille de la baie. La diversité nucléotidique de la région a

également été étudiée chez l'Ugni Blanc mutant afin de développer de nouveaux marqueurs pour encadrer la mutation et participer à la cartographie fine du gène en collaboration avec l'UMR AGAP (Montpellier) et le CSIRO (Adélaïde, Australie).

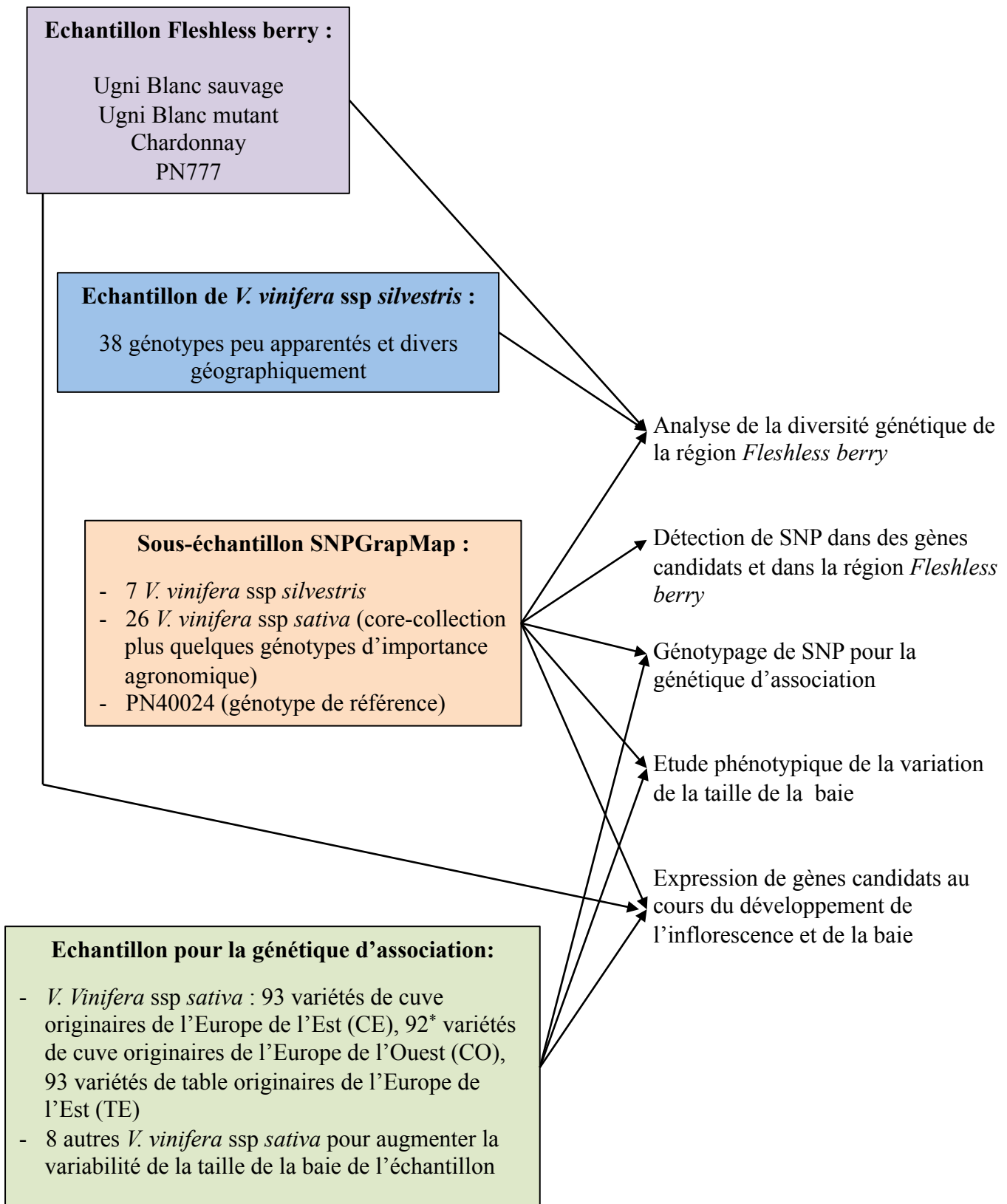
Ces trois aspects de mon travail sont présentés après le matériel et méthodes, sous la forme de trois chapitres, comprenant chacun une introduction, des résultats sous la forme d'un article, éventuellement des résultats complémentaires, et une conclusion. Enfin, une dernière partie présente la discussion générale et les perspectives de mon travail de thèse.



## 2. Matériel et méthodes



*Photographie d'une inflorescence au stade nouaison*



\* Un géotype a été retiré car il appartenait au compartiment sauvage.

**Figure 2.1:** Matériel végétal utilisé au cours de ma thèse.

## 2.1. Matériel végétal

L'ensemble des génotypes utilisés au cours de ma thèse provient du domaine expérimental de Vassal (INRA, Montpellier), conservatoire le plus important du monde pour la vigne cultivée *Vitis vinifera* L. Les seules exceptions sont le génotype de référence PN40024 (Jaillon et al, 2007) qui provient d'une lignée quasi homozygote maintenue à l'INRA de Colmar et le génotype Pinot Noir clone ENTAV-INRA777 (PN777) qui est conservé à l'Institut Français de la Vigne et du Vin (IFVV, Grau du Roi, France). Les génotypes, ainsi que leurs utilisations au cours de ma thèse, sont présentés ci-après (Figure 2.1 et Figure 1.3) et listés dans les annexes 1, 3 et 4-B. Ils ont été choisis par l'équipe DAVEM (UMR AGAP, Montpellier), sur la base des connaissances accumulées sur la diversité génétique chez les *Vitaceae*.

### Echantillon SNPGrpMap pour la découverte de SNP dans le genre *Vitis*

L'ensemble des génotypes *V. vinifera* de l'échantillon SNPGrpMap a été utilisé, soit:

- 7 génotypes *V. vinifera* ssp *silvestris* (sauvages): les moins apparentés parmi les génotypes sauvages conservés au domaine de Vassal, sur la base d'une étude à l'aide de 20 marqueurs SSR (Laucoy et al 2011);
- 27 génotypes *V. vinifera* ssp *sativa* (cultivés): 23 de la core-collection maximisant la diversité génétique de la vigne cultivée (Le Cunff et al, 2008), le génotype de référence PN40024 (Jaillon et al, 2007), et trois autres génotypes d'intérêt agronomique et/ou utilisés comme parents dans des populations de cartographie génétique.

Cet échantillon a été utilisé pour la découverte de SNP dans des gènes candidats pour la taille de la baie et dans la région du gène *Fleshless berry*, mais également pour rechercher des traces de domestication dans cette dernière région. Enfin, l'ensemble des génotypes cultivés et un des sauvages (Lambrusque de Grésigny 1) ont été phénotypés pour différents caractères de l'ovaire et de la baie dans le cadre de l'étude phénotypique de la variation de la taille de la baie.

### Echantillon de variétés cultivées pour la génétique d'association

Pour réaliser des études de génétique d'association, il fallait disposer d'un échantillon plus large de génotypes. Pour cela, un autre échantillon de la collection du domaine de Vassal a été développé par l'équipe DAVEM:



- 278 génotypes peu apparentés, structurés en 3 sous-populations avec un minimum de 80% d'appartenance à l'une des populations, et représentant 80% de la diversité génétique en SSR de la vigne cultivée, soit plus de 90% de la diversité génétique en SNP (S. Nicolas, T. Lacombe et L. Le Cunff communication personnelle). Cet échantillon est composé de 93 variétés de cuve originaires de l'Europe de l'Est (CE), de 92 variétés de cuve originaires de l'Europe de l'Ouest (CO) et de 93 variétés de table originaires de l'Europe de l'Est (TE).
- 8 autres *V. vinifera* cultivées ont été ajoutées afin d'augmenter la variabilité de la taille de la baie dans l'échantillon.

J'ai phénotypé l'ensemble de ces variétés pour différents caractères de l'ovaire et de la baie et je les ai génotypés avec un ensemble de SNP identifiés dans le sous-échantillon SNPGrpMap.

### **Echantillon de 38 *V. vinifera* spp *silvestris***

38 génotypes de *V. vinifera* sauvages de différentes provenances géographiques (France, Allemagne, Maroc, Tunisie, Slovaquie, Algérie, Suisse), et suffisamment âgés pour pouvoir être prélevés (feuilles), ont été choisis par l'équipe DAVEM (R. Bacilieri, S. Nicolas, T. Lacombe et J-M. Boursiquot) pour être génotypés à l'aide de marqueurs SNP de la région *Fleshless berry*, identifiés dans l'échantillon SNPGrpMap afin de confirmer les résultats obtenus précédemment à partir des 7 sauvages du sous-échantillon SNPGrpMap.

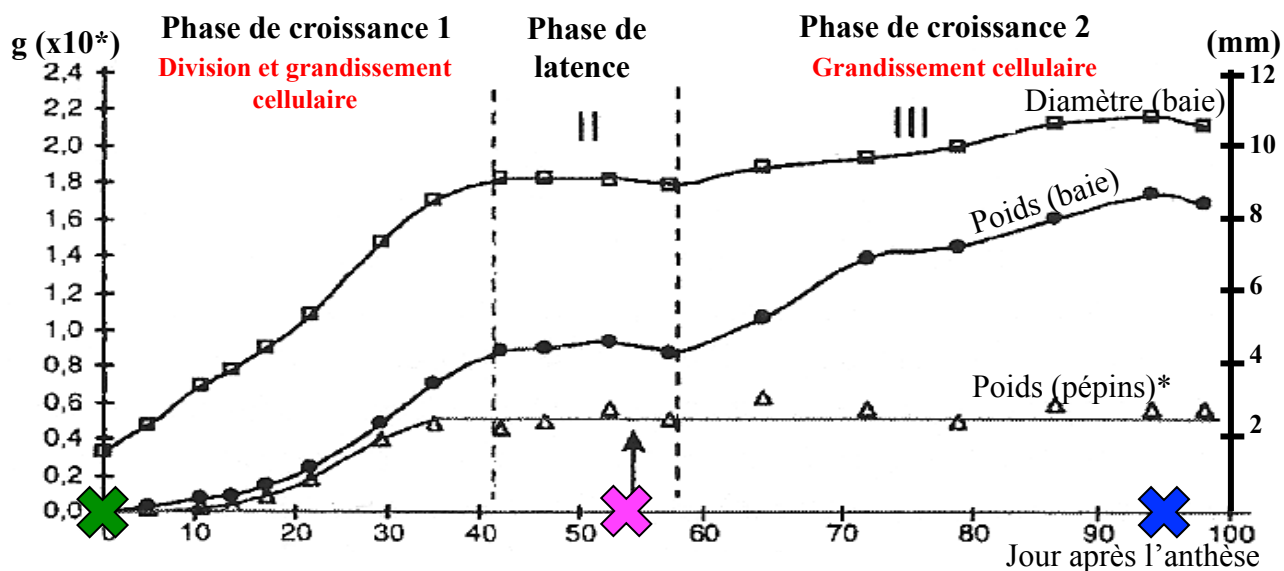
### **Echantillon Fleshless Berry**

4 autres génotypes *V. vinifera* cultivés ont également été étudiés dans le cadre de la recherche de polymorphisme dans la région du gène *Flb*:

- les parents de la descendance ayant servi à l'établissement de la carte génétique de la mutation *Fleshless berry* (Fernandez et al, 2006c), soit Ugni Blanc mutant et Chardonnay;
- l'Ugni Blanc (non muté au locus *Flb*);
- le Pinot Noir clone ENTAV-INRA777.

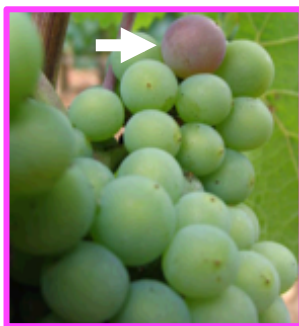
L'ensemble de ces génotypes a également été utilisé dans le cadre d'expériences de PCR quantitative sur ADN génomique et certains dans des expériences d'hybridation fluorescente *in situ* (FISH).

Enfin, deux génotypes très contrastés pour la taille de la baie, appartenant à l'échantillon de



Sur 10 ovaires:

- Diamètre
- Nombre de carpelles



Sur 30 baies:

- Poids

Sur 10 baies:

- Diamètre
- Nombre de pépins
- Poids des pépins



Sur 30 baies:

- Poids

Sur 10 baies:

- Diamètre

**Figure 2.2:** Stades du développement ciblés (d'après Ojeda et al, 1999) pour le phénotypage macroscopique des ovaires et des baies. La flèche blanche indique une baie vérée, les autres baies vont bientôt vérier et contiennent le nombre final de cellules de la baie, le stade suivant (III) correspondant à une croissance des baies uniquement par une augmentation du volume cellulaire .

génotypes cultivés pour la génétique d'association, ainsi que l'Ugni Blanc sauvage et mutant ont été utilisés dans le cadre d'expériences de PCR quantitative en temps réel sur ADN complémentaire, afin de tester l'expression de gènes candidats au cours du développement de l'inflorescence et de la baie.

## **2.2. Méthodes**

### **2.2.1. Phénotypage pour l'étude des déterminants de la variation de la taille de la baie**

#### **2.2.1.1. Analyse macroscopique des ovaires et des baies de vigne**

Pour chaque génotype, les ovaires et les baies ont été phénotypés à différents stades du développement:

- les ovaires ont été récoltés juste avant l'anthèse, en récoltant des fleurs fermées sur une inflorescence contenant quelques fleurs ouvertes (Figure 2.2), les ovaires récoltés sont donc entièrement formés et non fécondés;
- les baies ont été récoltées à deux stades du développement:
  - juste avant la véraison en récoltant des baies dures sur une grappe contenant au moins une baie à véraison visiblement identifiable par un changement de couleur (Figure 2.2) et palpable par un ramollissement de la baie. A ce stade, le nombre de cellules finales de la baie est fixé (Figure 2.2) et les pépins sont entièrement formés et facilement séparables de la pulpe. Les pépins ont donc été phénotypés à ce stade;
  - à maturité soit 40 jours après la première récolte, lorsque les baies ont atteint leur taille maximale.

Plusieurs paramètres ont été mesurés sur l'ensemble des génotypes (Figure 2.2):

- le nombre moyen de carpelles par ovaire comptés à la loupe binoculaire à partir de 10 ovaires coupés en deux (Figure 2.3);
- le poids moyen de 30 baies coupées à la base du pédicelle et récoltées aléatoirement sur au moins deux grappes;
- le diamètre moyen de 10 baies choisies aléatoirement parmi les 30 et le diamètre moyen de 10 ovaires récoltés sur au moins deux inflorescences et débarrassés en partie des autres pièces florales à l'aide de pinces (Figure 2.3). La longueur (l) et la largeur





**Figure 2.3:** Phénotypage des ovaires et du diamètre des baies. Photos (A) à la loupe binoculaire de 10 fleurs épluchées pour la mesure du diamètre des ovaires, (B) d'ovaires coupés en deux pour le comptage du nombre de carpelles et (C) de 10 baies juste avant le stade véraison à gauche et à maturité à droite pour la mesure du diamètre des baies.



(L) moyenne ont été mesurées à partir de photographies grâce au logiciel ImageJ® version 1.43 pour macintosh (<http://rsbweb.nih.gov/ij/download.html>);

- le nombre et le poids moyens de pépins par baie mesurés sur 10 baies.

Plusieurs paramètres ont été estimés à partir de ces mesures:

- le poids moyen de la pulpe d'une baie a été estimé en soustrayant le poids des pépins au poids de la baie;
- les formes de l'ovaire et de la baie ont été estimées à partir du ratio longueur/largeur. Si la valeur de ce ratio était égale à 1, l'ovaire ou la baie était alors déclaré(e) rond(e), alors qu'une valeur >1 témoigne d'une forme allongée;
- les volumes de l'ovaire et de la baie ont été approchés comme le volume d'un ellipsoïde en utilisant la formule  $V=4/3.\pi.L.L^2$ .

### 2.2.1.1. Analyse microscopique des ovaires et des baies de vigne

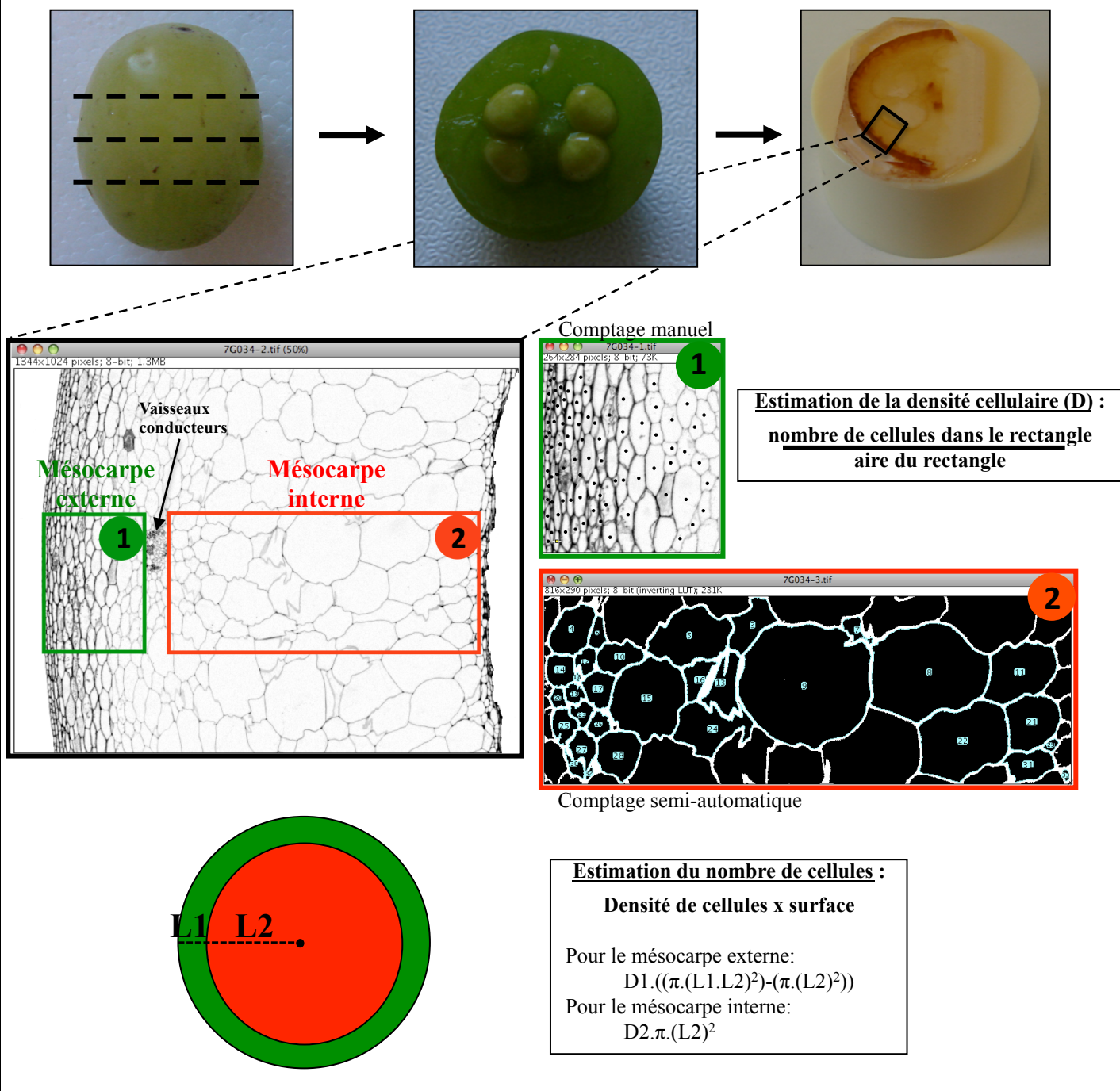
#### 2.2.1.1.1. Préparation des échantillons

L'analyse microscopique a été réalisée sur 26 génotypes représentatifs des 3 sous-populations de l'échantillon pour la génétique d'association et sur un génotype *V. vinifera* sauvage (Annexe 1). Trois ovaires et trois baies par génotype ont été immergés immédiatement après avoir été cueillis dans un fixateur glutaraldéhyde/paraformaldéhyde/caféine, puis placés sous vide 3 fois 10 minutes (min) pour faire pénétrer le fixateur. Ce fixateur a ensuite été remplacé par une nouvelle solution de fixateur, et les échantillons ont été conservés en chambre froide pendant au moins 48 heures (h). A cette étape, les échantillons peuvent rester plusieurs mois dans le fixateur.

#### Composition du fixateur:

- Tampon phosphate de sodium 0,1 M pH 7, 2	50 mL
- Paraformaldéhyde à 10%	20 mL
- Glutaraldéhyde à 10%	10 mL
- H <sub>2</sub> O	19 mL
- Caféine	1g

Après fixation, les échantillons ont été rincés dans une solution d'éthanol 70%, les baies ont alors été épépinées et coupées en tranches à l'aide d'un scalpel, seules les deux tranches du milieu avec le plus grand diamètre ont été gardées (Figure 2.2). Les échantillons ont été déshydratés en suivant une série d'incubation dans différentes concentrations d'éthanol: éthanol 70% 1 h, rinçage à l'éthanol 95%, éthanol 95% 1 h, rinçage à l'éthanol 100%, et



**Figure 2.4:** Estimation du nombre et du volume des cellules des baies à véraison. Comptage semi-automatique ou manuel du nombre de cellules du mésocarpe grâce au logiciel ImageJ® et estimation du nombre de cellules et de la densité cellulaire sur une coupe transversale de baie.

éthanol 100% 24 h à 4°C. Les échantillons ont ensuite été incubés dans un milieu d'imprégnation et conservés à l'obscurité en chambre froide au minimum pendant 15 jours dans cette solution.

Composition du milieu d'imprégnation:

- Résine Kulzer	500 mL
- Initiateur	5 g
- Polyéthylène glycol 400	10 mL

Chaque échantillon a été placé dans un moule, dans lequel a été ajouté la résine composée du milieu d'imprégnation et du polymérisateur (1 mL d'éthylène glycol diméthacrylate pour 15 mL de milieu). Une fois la polymérisation amorcée, les échantillons ont été placés dans une étuve pendant 24 h. Les inclusions ont été collées sur un support et les échantillons ont été coupés en tranche de 2-3  $\mu\text{m}$  à l'aide d'un microtome (LEICA<sup>TM</sup> RM2255). Les coupes ont ensuite été déposées sur lames de verre, et chauffées sur une platine à 50°C durant 10 minutes. Les lames contenant les coupes ont alors été immergées dans la solution colorante (bleu Alcian 8 GX 0,1%, citrate de sodium 0,01 M, HCl 0,01 M pH 3,5), et conservées à l'obscurité et à température ambiante pendant 24 h. Après coloration, les lames ont été rincées 3 fois dans un tampon citrate de sodium 0,01 M, HCl 0,01 M pH 3,5, et chauffées sur une platine à 50°C durant 10 min. Enfin, les coupes ont été montées sous hotte entre lame et lamelle en utilisant le milieu de montage à base de xylène (Isomount medium; Labonord) et laissées à sécher plusieurs heures.

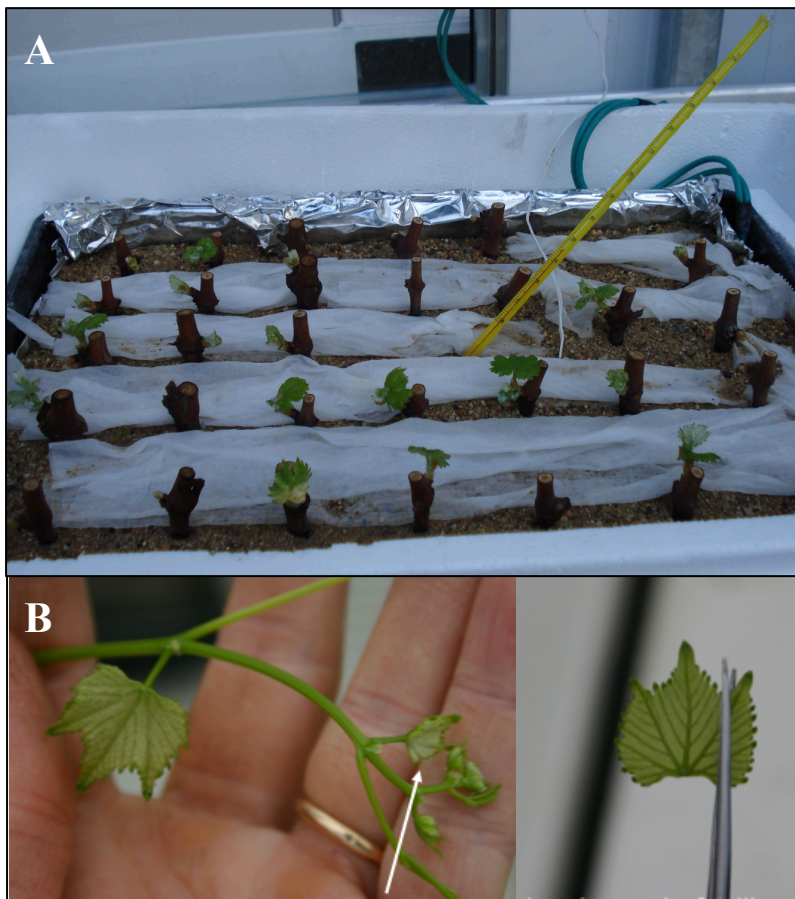
#### **2.2.1.1.2. Analyse microscopique**

Les images des coupes d'ovaires et de baies ont été acquises à l'aide d'un microscope (DM 4500, Leica<sup>TM</sup>) couplé à une caméra numérique (Qimaging; Micropublisher 3.3 RTV) et piloté par le logiciel d'analyse d'images Velocity® version 5.4.1 pour macintosh (<http://www.cellularimaging.com/products/velocity/>). Trois objectifs ont été utilisés:

- x5 pour mesurer la longueur du mésocarpe interne et externe des baies et pour le comptage du nombre de cellules des baies (Figure 2.4);
- x10 pour mesurer la longueur du mésocarpe interne et externe des ovaires;
- x20 pour le comptage du nombre de cellules de l'ovaire.

La meilleure image de chaque ovaire et baie a été choisie, et si besoin reconstruite par fusion des images grâce au logiciel Photoshop® version CS3.

Le nombre de cellules a été compté grâce au logiciel ImageJ® de deux manières différentes (Figure 2.4):



**Figure 2.5:** Matériel végétal récolté pour les expériences de BAC-FISH. **(A)** Dispositif des cultures un oeil **(B)** très jeunes feuilles ou bordures de jeunes feuilles récoltées (photo A. Canaguier et M-S. Verneret)

- manuellement pour le mésocarpe externe des ovaires et baies et le mésocarpe interne des ovaires;
- semi-automatiquement pour le mésocarpe interne des baies, rendu possible grâce au fait que les cellules du mésocarpe interne ont une paroi cellulaire constante et fine et par la quasi absence de bruit de fond.

La densité cellulaire du mésocarpe interne et externe (facilement identifiable car séparés par les vaisseaux conducteurs dans l’ovaire et la baie; Figure 2.4), a été estimée comme le nombre de cellules entières dans une aire définie (Figure 2 4). La densité cellulaire étant inversement proportionnelle au volume cellulaire, une densité cellulaire élevée reflète un nombre important de cellules par unité de surface donc de petites cellules et inversement. La densité cellulaire a donc été utilisée par la suite, comme estimateur de la taille des cellules.

Le nombre total de cellules présentes dans la section du mésocarpe a été estimé pour les mésocarpes interne et externe en tenant compte de leurs proportions respectives comme étant égal à la densité de cellules multipliée par la surface du mésocarpe correspondant (Figure 2.4).

## **2.2.2. Préparation du matériel pour les expériences d’hybridation fluorescente *in situ* de clones BAC (BAC-FISH)**

### **2.2.2.1. Choix du matériel végétal**

Des boutures dites « un œil » (bois d’environ 10 cm contenant un bourgeon) ont été réalisées pour l’Ugni Blanc mutant, Ugni Blanc sauvage et le PN777 dans un sol contenant: 1/3 de perlite, 1/3 de gros sable et 1/3 de terreau. Ce mélange a été maintenu à une température de 25-28°C à l’aide d’une résistance (Figure 2.5). Les boutures ont été arrosées avec de l’eau tous les deux jours. Les très jeunes feuilles, les bordures de jeunes feuilles (Figure 2.5) ainsi que les extrémités des racines (environ 5 mm) ont été récoltées et immédiatement plongées dans une solution de 8-hydroxyquinoline à 0,04% (40 mg de 8-hydroxyquinoline dans 100 mL d’éthanol 96%) pendant 2 h à température ambiante, puis 2 h à 4°C. Le matériel végétal a ensuite été fixé dans une solution d’éthanol 100% et d’acide acétique glacial (3/1, V/V) préparée le jour même, pendant 24 h à 4°C, puis rincé et conservé dans l’éthanol 70% à 4°C avant l’envoi à la plateforme de cytologie de l’INRA de Rennes (O. Coriton, M-S Vernerey).

### 2.2.2.2. Choix des sondes

Les sondes utilisées pour les expériences de FISH étaient des BAC (*Bacterial Artificial Chromosomes*) provenant d'une banque réalisée avec le génotype PN40024 (Jaillon et al, 2007). Les séquences d'extrémités de ces BAC ont été alignées par l'URGI (INRA, Versailles) sur la séquence de référence de la vigne et ces alignements sont visualisables sur leur site d'accès aux données du séquençage ([http://urgi.versailles.inra.fr/cgi-bin/gbrowse/vitis\\_12x\\_pub/](http://urgi.versailles.inra.fr/cgi-bin/gbrowse/vitis_12x_pub/)). Deux BAC ont été choisis dans la région *Fleshless berry* selon les critères suivants: ils devaient couvrir une région avec peu d'éléments répétés, se trouver dans une région ne présentant pas de duplication récente sur le génome et ils devaient être non chevauchants entre eux.

Identifiant du BAC	Position sur le chromosome 18
VV40024H140P14	861381.. 944405
VV40024H040I16	684275.. 773827

Les BAC sélectionnés ont été extraits à l'aide du protocole suivant. Un tube contenant 1 mL de milieu FM glycérol 6% stérile (30 mL de milieu FM glycérol 6%, 300 µl de MgSO<sub>4</sub> à 40 mM et 11,25 µl de chloramphenicol à 50 mg/mL) a été inoculé sous hotte à flux laminaire en conditions stériles à l'aide d'un cure-dent à partir d'une copie de la banque BAC conservée au laboratoire. Les clones BAC ont été laissés à se multiplier à 37°C pendant 24 h sous agitation. 1,5 µl des cultures ainsi obtenues ont ensuite été déposés dans un tube à vis contenant 1 mL de milieu LB-agar-chloramphénicol (2,25 µl de chloramphénicol à 50 mg/mL pour 6 mL de milieu LB-agar) et laissés pousser à 37°C pendant 24 h, puis conservés à 4°C avant l'envoi à la plateforme de cytologie de l'INRA de Rennes.

#### Composition du milieu FM 6 % glycérol:

- LB	20 g
- KH <sub>2</sub> PO <sub>4</sub> anhydre	1,42 g
- K <sub>2</sub> HPO <sub>4</sub> anhydre	5,02 g
- Sodium citrate	4 g
- Ammonium Sulfate	0,72 g
- Glycerol 100%	48 g
- H <sub>2</sub> O	qsp 800 mL

#### Composition du milieu LB-agar:

- Bactotrytone	1 g
- Extrait de levure	0,5 g
- NaCl	1 g
- Agar	1, 5 g
- H <sub>2</sub> O	qsp 100 mL

### 2.2.3. Extraction et quantification d'ADN génomique

Les ADN de l'échantillon SNPGrpMap utilisés dans les expériences de séquençage nous ont été fournis par l'équipe DAVEM de l'INRA de Montpellier.

Chaque extraction réalisée au laboratoire a été faite en triplicat sur l'échantillon de variétés cultivées pour la génétique d'association, l'échantillon de 38 *V. vinifera* sauvages et l'échantillon *Fleshless berry*.

#### 2.2.3.1. A partir de feuilles

Un disque foliaire de 50 mg a été coupé à l'emporte pièce puis déposé dans les puits d'une plaque 96 contenant une bille d'acier de 4 mm de diamètre posée sur un bloc de glace le temps de la récolte. Une deuxième bille d'acier a été ajoutée avant la fermeture des plaques, qui ont ensuite été conservées à -80°C avant transfert dans de la carboglace de Montpellier à Evry. Pour le broyage des disques foliaires, les plaques 96 ont été posées dans un bac contenant de l'azote liquide, puis agitées à l'aide d'un mélangeur de peinture. L'ADN génomique a ensuite été extrait en utilisant un kit Quiagen (DNeasy 96 Plant Kit) et en suivant les instructions du fournisseur, puis conservé à -20°C.

#### 2.2.3.2. A partir de bois

Le bois peut être conservé plusieurs jours à température ambiante, permettant le transfert de Montpellier à Evry par la poste. L'écorce du bois a été retirée à l'aide d'un scalpel et le liber restant a été coupé en rondelles de 2-4 mm d'épaisseur à l'aide d'un sécateur. Les rondelles ont ensuite été broyées dans l'azote liquide à l'aide d'un mortier et d'un pilon. L'ADN génomique a été extrait à partir d'1 g de poudre en utilisant un kit Qiagen (DNeasy Plant Maxi Kit) en suivant les instructions du fournisseur, puis conservé à -20°C.





### 2.2.2.3. A partir de baies

Dix baies par génotype ont été coupées en quatre, épépinées puis plongées dans l'azote liquide. Les baies ont été broyées à l'aide d'un broyeur à billes de type « Dangoumau » ne prenant qu'un échantillon à la fois (disponible dans l'équipe DAVEM). 2 g de poudre de baies ont été transférés dans un tube de 50 mL auxquels on a ajouté 25 mL de tampon de lyse I à 4°C sous hôte chimique puis agités et laissés dans la glace, le temps de préparer les autres tubes. Les tubes ont ensuite été centrifugés à 3500 g à 4°C et le surnageant a été éliminé. A cette étape, les culots peuvent être stockés à -20°C pendant plusieurs jours (permettant le transfert de Montpellier à Evry dans de la carboglace).

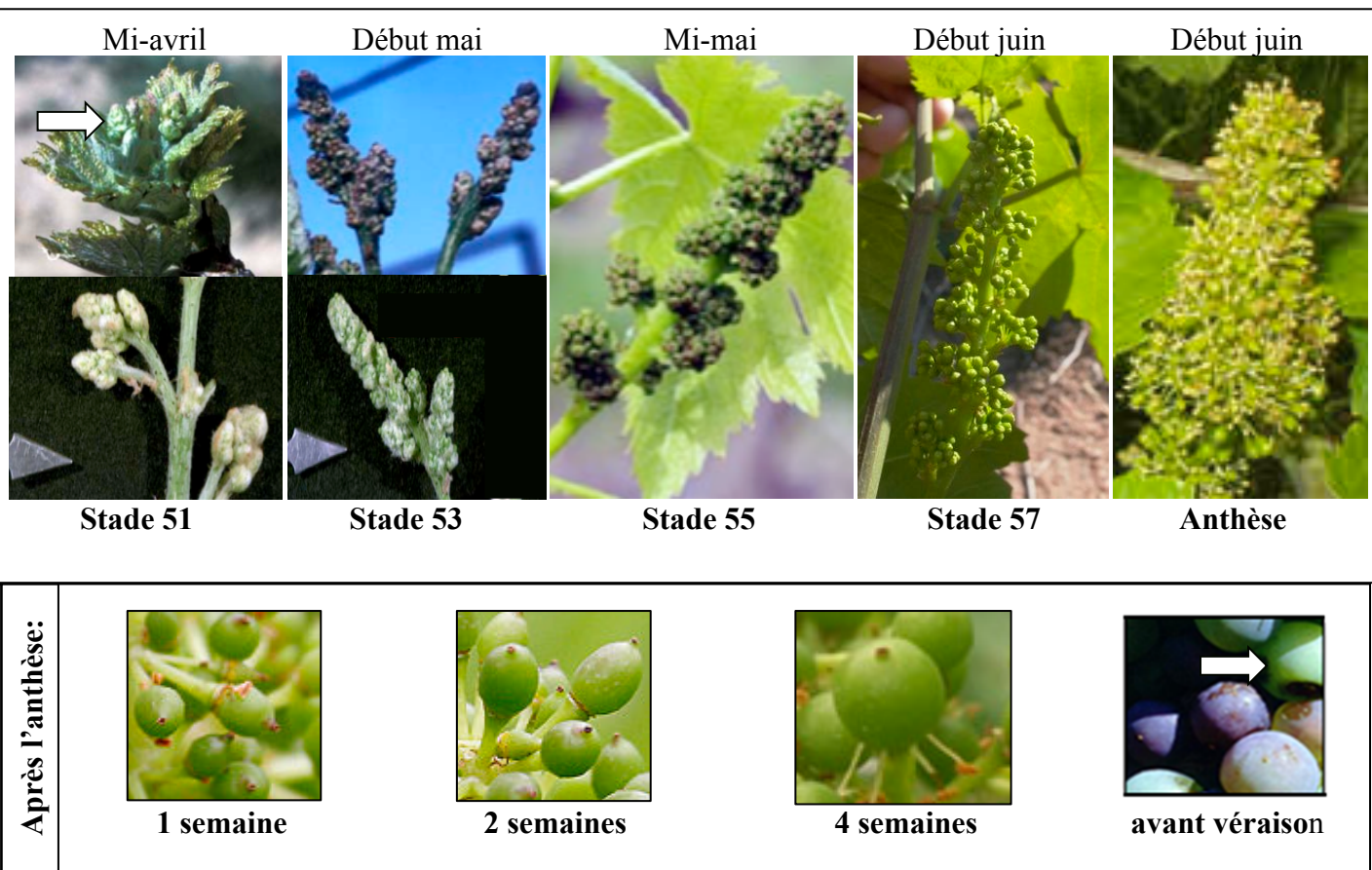
#### Composition du tampon de lyse I:

- TrisHCl 1 M pH 8	200 mL
- Na <sub>2</sub> EDTA 0,5 M pH 8	100 mL
- NaCl 5 M	50 mL
- Polyvinylpyrrolidone 40 000	25 g
- βmercaptoéthanol	10 mL (ajouter juste avant l'emploi)
- H <sub>2</sub> O	qsp 1 L

Les étapes suivantes ont été réalisées sous une hotte chimique. Le culot a été repris dans 20 mL de tampon de lyse II, puis placé au bain-marie sous agitation 4 h à 65°C. 20 mL de chloroforme/isoamylique (24/1 v/v) ont été ajoutés à la solution et les tubes ont été agités par retournement puis centrifugés 10 min à 4000 g à température ambiante. La phase aqueuse a été transférée dans un tube de 13 mL et 25 µL de Ribonucléase A à 10 mg/mL (Qiagen) ont été ajoutés. Les tubes ont alors été mis à incuber 30 min dans une étuve à 37°C. 20 mL d'isopropanol à 4°C ont été alors ajoutés pour précipiter l'ADN et les tubes ont été agités par retournement puis centrifugés à 3000 g pendant 5 min. Le culot a été repris dans 0,7 mL de tampon TE 10:1 pH 8 (Tris HCl 10 mM et Na<sub>2</sub>EDTA 1 mM) et le tout a été transféré dans un tube 2 mL. 70 µL d'acétate d'ammonium à 3 M pH 5, 2 et 0, 6 mL d'isopropanol ont alors été ajoutés et le tube a été centrifugé 2 min à 14000 g. Le culot a été rincé à l'aide de 1 mL d'éthanol 70% et centrifugé 2 min à 14000 g. L'éthanol a été retiré puis les tubes ont été centrifugés à vide 1 min à 14000 g puis placés à 37°C à l'étuve 10-15 min pour sécher le culot. Enfin, le culot a été repris dans 100 µL de tampon TE 10:1 pH 8 et conservé à -20°C.

#### Composition du tampon de lyse II:

- Tris HCl 1 M pH 8	200 mL
- Na <sub>2</sub> EDTA 0,5 M pH 8	100 mL



**Figure 2.6:** Stades de développement auxquels ont été faits les prélèvements d'ovaires et de baies pour les extractions d'ARN (stades du code international BBCH; Lorenz et al, 1994; photos L. Torregrosa et C. Houel).

- NaCl	73 g
- Alkyltrimethyl ammonium bromide	40 g
- $\beta$ mercaptoéthanol	10 mL (ajouter juste avant l'emploi)
- H <sub>2</sub> O	qsp 1 L

#### 2.2.3.4. Quantification de l'ADN et vérification de sa qualité

Pour toutes les extractions (feuilles, bois, et baies) 2  $\mu$ L d'ADN, 2  $\mu$ L de bleu de charge 6X et 6  $\mu$ L d'eau stérile ont été déposés et mis à migrer (Voltage constant, 300 mA, 35 min) sur gel d'agarose 1 % dans du tampon TAE 1X afin de vérifier la qualité des ADN et la digestion des ARN.

Les ADN génomiques extraits ont ensuite été quantifiés en plaque 96 noire en utilisant le kit Invitrogen (Quant-iT™ PicoGreen® dsDNA Assay Kit). Dans chaque puits ont été déposés 50  $\mu$ L de l'ADN dilué au 1/150 dans du TE 10:1 pH 8 et 50  $\mu$ L de PicoGreen dilué au 1/200 dans du TE 10:1 pH 8. Une gamme étalon d'ADN de sperme de saumon (Applied Biosystems, 50 ng/ $\mu$ L) a également été réalisée (les 5 points de la gamme sont: 0,5 ng, 1,5 ng, 3 ng, 4 ng, 5 ng afin d'estimer la concentration des ADN extraits. Après remplissage, les plaques à doser ont été agitées pendant 5 min, puis analysées grâce au logiciel Biolise (version 2.6.5 Rev 7).

### 2.2.4. Extraction d'ARN et transcription inverse

#### 2.2.4.1. Extraction ARN

Des inflorescences à différents stades du développement (stade 51, stade 53, stade 55, stade 57 (stades du code international BBCH; Lorenz et al, 1994) et à l'anthèse; Figure 2.6), ainsi que des baies en formation à différents stades après l'anthèse (1 semaine, deux semaines et quatre semaines après l'anthèse) ont été récoltées pendant deux années consécutives (2009 et 2010) et immédiatement plongées dans l'azote liquide, puis conservées à -80°C. Les inflorescences ont été broyées dans l'azote liquide dans un tube de 1,5 mL à l'aide d'un goupillon, alors que les baies en formation ont été broyées dans l'azote liquide, au mortier et au pilon. Les Acides RiboNucléiques (ARN) totaux ont été extraits à partir de 50 mg de poudre en utilisant un kit Sigma (Spectrum™ Plant total RNA Kit 50), en suivant les instructions du fournisseur y compris l'étape de dégradation de l'ADN par la Desoxyribonucléase I (Sigma), puis conservés à -80°C.

### 2.2.3.2. Quantification et transcription inverse

Après homogénéisation, des dilutions au 1/10 dans de l'eau stérile des ARN totaux ont été réalisées. 1 µL de chaque dilution a été déposé dans une plaque 96 noire, avec 49 µL de TE 10:1 pH 8 stérile et 50 µL de RiboGreen® (Invitrogen) au 1/200. Pour quantifier les ARN totaux, une gamme étalon de 0,5-10 kb marqueur ARN (Invitrogen) a été réalisée (à l'aide de 8 points de la gamme: 2,5 ng, 5 ng, 10 ng, 20 ng, 40 ng, 60 ng, 80 ng et 100 ng).

Tous les échantillons ont été ajustés à une concentration finale de 117,6 ng/µL. 1 µg d'ARN totaux (8,5µL), et 8,5µL de bleu de charge 2X (900 µL de formamide, 72 µL d'EDTA 0,5 M stérile et 1 µL de bleu de Xylène Cyanol), ainsi que le marqueur de taille ont été dénaturés (10 min à 95°C puis 5 min dans la glace) puis déposés sur un gel d'agarose 0,8% dans du tampon TAE 1X RNase free (migration de 30 min à 200 V), afin de vérifier leur qualité et l'homogénéité de leurs concentrations. La digestion de l'ADN a été vérifiée par PCR sur 1 µL d'ARN totaux avec le marqueur GAPDH (la taille attendue de l'amplicon est de 475 pb sur ADNg et 70 pb sur ARN; protocole de PCR paragraphe 2.2.5.2).

La transcription inverse des ARN messagers (RT) en ADN complémentaires (ADNc) a été réalisée en plaque 384 sur l'ensemble des échantillons dans un volume final de 20 µL à partir de 1 µg ARN total, de 1 µg d'amorces oligodT (12-18 mer; Invitrogen), 10 mM de dNTP (Applied Biosystem). La plaque a été chauffée 10 min à 65°C à l'aide d'une machine PCR, puis déposée dans la glace, après ajout de 4 µL de tampon 5X et 200 µmol de DiThioThréitol (DTT), la plaque a été chauffée 5 min à 42°C puis déposée dans la glace, 200 unités de transcriptase inverse Superscript II (Invitrogen) ont été ajoutés et la plaque a de nouveau été chauffée à 50 min à 42°C puis déposée dans la glace. Un traitement avec 1 unité de Ribonucléase H (Invitrogen) 20 min à 37°C a été appliqué à l'ensemble des échantillons. Les témoins négatifs de l'ensemble des échantillons (sans transcriptase inverse ou RT-) ont été réalisés et traités à la Ribonucléase H (Invitrogen).

## 2.2.5. PCR

### 2.2.5.1. Définition des couples d'amorces

Tous les couples d'amorces utilisés au cours de ma thèse ont été définis grâce au logiciel Primer 3 version 0.4.0 (<http://frodo.wi.mit.edu/primer3/>; Rozen et Skaletsky, 2000). Au début de ma thèse l'annotation de la version 8X d'assemblage du génome de référence PN40024 a été utilisée (Jaillon et al, 2007), puis j'ai revu les résultats par rapport à la nouvelle version 12X d'assemblage du génome. Ces deux annotations sont disponibles sur le site d'accès aux

données du séquençage de la vigne du Génoscope (<http://www.genoscope.cns.fr/externe/GenomeBrowser/Vitis/>), et dans la base de données FlagDB++ (<http://urgv.evry.inra.fr/projects/FLAGdb++/HTML/index.shtm>).

- Pour la recherche de polymorphisme de séquence dans la région *Fleshless berry*, un couple d'amorces a été défini pour chaque gène en respectant les critères suivants: longueur de l'amplicon 1000-1400 pb, température d'hybridation (T<sub>m</sub>) comprise entre 53-57°C avec un optimum à 55°C, amorces positionnées sur les exons de part et d'autre d'un intron afin de détecter le maximum de polymorphisme et d'amplifier le maximum de génotypes car les séquences exoniques sont mieux conservées d'un génotype à l'autre;
- Pour les gènes candidats pour la variation de la taille de la baie, plusieurs couples d'amorces ont été définis permettant de couvrir la partie codante du gène et si possible également l'ensemble des introns. La couverture des introns s'est avérée souvent difficile car ils sont très polymorphes d'un génotype à l'autre avec beaucoup d'insertions/délétions à l'état hétérozygote rendant les séquences Sanger des amplicons illisibles. Les couples d'amorces ont été définis selon les critères suivants: température d'hybridation (T<sub>m</sub>) comprise entre 53-57°C avec un optimum à 55°C, amorces sur exons ou introns/UTR (à une distance <150 bp d'un exon car la séquence est encore bien conservée d'un génotype à l'autre; M. Moroldo communication personnelle) afin d'amplifier le maximum de séquences exoniques.

Des séquences correspondant aux amorces universelles SP6 (pour l'amorce sens) et T7 (pour l'amorce anti-sens) ont ensuite été rajoutées aux séquences des amorces ainsi définies pour faciliter le séquençage de ces nombreux fragments.

- Afin d'étudier l'expression de gènes candidats au cours du développement de la fleur et de la baie, trois couples d'amorces par gène ont été définis sur les critères suivants: longueur de l'amplicon 60-150pb (sur ADNc), température d'hybridation (T<sub>m</sub>) comprise entre 53-57°C avec un optimum à 55°C, amorces dans le dernier exon (région plus conservée, L. Torregrosa communication personnelle) ou encadrant le dernier intron en s'appuyant sur les données d'ADN complémentaire et d'EST. Neuf couples d'amorces correspondant à des gènes de référence constitutivement exprimés au cours du développement de la fleur et de la baie déjà publiés (Reid et al, 2006; Fernandez et al, 2007) ont été analysés (Annexe 2) afin de choisir les gènes présentant un profil d'expression le plus stable possible au cours des différents stades testés.

### 2.2.5.2. Test génomique

Chaque couple d'amorces a été testé par PCR sur le PN40024 (témoin positif), deux *V. vinifera* (ssp *sativa* Buckland Sweet Water et ssp *sylvestris* Oberlin 21 mâles pour les couples d'amorces de gènes de la région *Flesless berry* et pour les gènes candidats et sur les cv. Ugni blanc sauvage et Ugni blanc mutant pour les couples d'amorces des expériences de PCR quantitative en Temps-Réel (q-RT-PCR)) et l'eau (témoin négatif) en suivant le programme suivant: 5 min à 94°C; 14 cycles de (20 secondes (s) à 94 °C, 20 s de 65 à 51°C (-1°C par cycle), 40 s à 72°C), puis 35 cycles de (20 s à 94°C, 20 s à 50 °C et 40 s à 72°C); et 7 min à 72°C. La totalité du produit PCR a été déposée sur un gel d'agarose Métaphor® 3% (Lonza) dans tampon du TEA 1% (migration à voltage variable, 300 mA, 50 min). Seuls les couples d'amorces donnant un profil simple bande pour les trois génotypes ont été gardés pour la suite des travaux.

#### Composition du mélange réactionnel de PCR:

- Tampon II 10X (Invitrogen)	1 µl
- MgCl <sub>2</sub> 25 mM	0,64 µl
- dNTP 10 mM	0,4 µl
- Solution de rouge de crésol	2 µl
(Rouge de Crésol 100 mg, saccharose 30 g et H <sub>2</sub> O qsp 50 mL)	
- Mélange d'amorces (sens et anti-sens) 12 pmol/µl	1 µl
- Taq polymérase du laboratoire 12X	0,1 µl
- ADN 2-4 ng/µl	2 µl
- H <sub>2</sub> O	qsp 10 µl

### 2.2.5.3. Séquençage

Les produits PCR utilisés pour la réaction de séquençage ont été synthétisés en plaque 96, à l'aide du programme suivant: 5 min à 94°C; puis 35 cycles de: 30 s à 94°C, 30 s à 55°C et 1 min à 68°C; et enfin 7 min à 68°C.

#### Composition du mélange réactionnel de PCR:

- Tampon 10X	2 µl
- MgSO <sub>4</sub> 50 mM	0,8 µl
- dNTP 10 mM	0,4 µl
- Mélange d'amorces (sens et anti-sens) 10 µM	0,8 µl
- Taq polymérase Platinum (Invitrogene)	0,15 µl

- |                    |           |
|--------------------|-----------|
| - ADN 2-4 ng/μl    | 2 μl      |
| - H <sub>2</sub> O | qsp 20 μl |

10 μl de chaque produit PCR mélangés avec 2 μl de bleu de charge 6X ont été déposés sur gel d'agarose 1% dans un tampon TAE 1% (migration à voltage variable, 300 mA, 50 min) afin de vérifier l'amplification du fragment. Les 10 μl restants ont ensuite été transmis à l'EPGV, équipe localisée sur la plateforme du Centre National de Génotypage (Evry), qui a réalisé les étapes de séquençage en Sanger dans le cadre du projet SNPGrMap.

#### 2.2.5.4. PCR quantitative

L'ensemble des expériences de PCR quantitative (q-PCR) a été réalisé par A. Canaguier et R. Berthomé (URGV, Evry).

##### 2.2.5.4.1. Sur ADN génomique

Afin de vérifier si l'Ugni Blanc mutant présentait une délétion à l'état hétérozygote de la région *Fleshless berry*, la quantité d'ADN d'un gène dans la région *Flb* en copie unique dans le génome a été mesurée par q-RT-PCR chez l'Ugni Blanc mutant et le PN777 utilisé comme contrôle (car il ne présente pas de délétion de la région). La quantité du produit d'amplification d'un gène unique dans la région *Flb* a été normalisée par la quantité du produit d'amplification d'une famille de trois gènes localisés en dehors de la région *Flb*. S'il y a une délétion hétérozygote de la région *Flb* chez l'Ugni Blanc mutant, la quantité du gène unique dans cette région après normalisation sera égale à ½, alors que s'il n'y a pas de délétion cette quantité sera égale à 1. Dans tous les cas cette quantité sera égale à 1 chez le PN777. Le protocole qui a été utilisé est le suivant.

Afin de quantifier l'ADN présent dans les échantillons analysés, une gamme étalon de quatre points (4 ng, 2 ng, 0,2 ng, et 0,02 ng) a été réalisée à partir de l'ADN du génotype de référence PN40024 et a été testée sur deux couples d'amorces. Les réactions de PCR ont été réalisées en plaques 384 dans un volume final de 15 μL par puits contenant: 5 pmol de chaque amorce, 2,5 μL d'ADN à 0,25 ng/μL et 7,5 μL du mélange de réactionnel SYBR Green® Master Mix 2X (Applied Biosystems) et en suivant les recommandations du fournisseur (10 min à 95°, suivies par 40 cycles de 15 s à 95°C et 1 min à 60°C). Chaque réaction de q-PCR a été réalisée en triplicat. La synthèse d'ADN double brin a été détectée par le système Applied Biosystems (ABI PRISM® 7900HT Sequence Detection System). L'efficacité de la PCR a été calculée pour chaque couple d'amorces et la courbe de dissociation de chaque amplicon permettant de visualiser la spécificité de chaque réaction PCR a été obtenue après le cycle





suivant: 15 s à 95°C, 15 s à 60°C, 15 s à 95°C. Ayant atteint le seuil de la phase exponentielle de la réaction PCR, la quantité d'ADN génomique correspondant au gène cible avec son écart-type ont été estimés en utilisant le logiciel SDS version 2.2.2 (Applied Biosystems). Les produits PCR ont ensuite été envoyés à séquencer (société Genoscreen), puis alignés (Blastn avec les paramètres par défaut) sur le génome de référence *via* sur le site d'accès aux données du séquençage de la vigne du Génoscope pour vérifier la conformité du produit PCR amplifié.

#### 2.2.5.4.2. Sur ADN complémentaire

Deux échantillons de référence ont été préparés: le premier à partir d'une dilution au 1/10 du mélange d'1 µL de chaque ADNc (tous génotypes, stades et années de récolte confondus) et a été utilisé comme échantillon de référence et témoin positif. Le second échantillon a été préparé à partir d'une dilution 1/10 du mélange d'1 µL de chaque produit RT- (voir paragraphe 2.2.3.2 ; tous génotypes, stades et années de récolte confondus) et a été utilisé comme le témoin négatif.

Chaque paire d'amorces a été testée sur une série de dilutions de l'échantillon de référence (1/20, 1/100, 1/1000, 1/5000) et sur la solution 1/10 du témoin négatif afin d'obtenir l'efficacité des amorces en PCR et de générer la courbe de dissociation de chaque amplicon. Trois triplicats techniques pour chaque couple d'amorces ont été inclus pour chaque réaction de q-RT-PCR. Les protocoles et programme suivants ont été utilisés: 2,5 µL d'ADNc de référence aux différentes dilutions, 5 µL du mélange d'amorces à 1 µM et 7,5 µL du mélange SYBRGreen PCR master mix 2X (Eurogentec), 10 min à 95°C, suivies par 40 cycles de 15 s à 95°C et 1 min à 60°C. La synthèse d'ADN double brin a été détectée par le système Applied Biosystems (ABI PRISM® 7900HT Sequence Detection System). La courbe de dissociation de chaque amplicon a été obtenue après le cycle suivant: 15 s à 95°C, 15 s à 60 °C, 15 s à 95°C permettant de tester la spécificité de chaque réaction PCR et l'efficacité de la PCR a été calculée pour chaque couple d'amorces. Les produits PCR ont ensuite été déposés sur un gel d'agarose Metaphor® 3% dans du tampon TAE 1X afin de vérifier la taille correcte du produit PCR et l'absence de bandes multiples. Pour chaque gène candidat, le couple d'amorces avec la meilleure efficacité PCR (variant entre 97 et 100%) et ne présentant pas d'amplification non-spécifique a été gardé. Pour les gènes de référence les mêmes critères ont été appliqués et l'expression constitutive du gène durant le développement de la fleur et de la baie vérifiée. La q-RT-PCR a été réalisée comme décrite précédemment avec le meilleur couple d'amorces pour chaque gène candidat et les deux meilleurs gènes de référence sur le mélange ADNc de référence (1/20, 1/100, 1/1000) et sur une dilution 1/100 des ADNc des

[illegible]Témoïn – (H<sub>2</sub>O)

**Figure 2.7:** Exemple d'alignement des séquences et de détection des SNP dans le sous-échantillon SNPGrpMap à l'aide du logiciel Genalys®.

différents stades de développement pour quatre génotypes et sur les deux années de récolte (répétitions biologiques) et en triplicats (répétitions techniques). Le seuil de la phase exponentielle de la réaction PCR (nombre de cycles d'amplification suffisant) étant atteint, la quantité d'ADNc de chaque gène candidat avec son écart-type ont été estimés, en utilisant le logiciel SDS version 2.2.2 (Applied Biosystems). L'expression moyenne des deux gènes de référence (*EFL-α* et *Aquaporin* correspondant respectivement aux gènes GSVIVG01025146001 et GSVIVG01016276001) à chaque stade, a été utilisée pour normaliser l'expression moyenne des gènes candidats aux stades respectifs (moyenne réalisée à partir des répétitions biologiques et techniques). Les produits PCR ont ensuite été déposés sur un gel d'agarose Metaphor® 3 % dans du tampon TAE 1 X afin de vérifier la taille correcte du produit PCR et l'absence de bandes multiples. Les produits PCR ont ensuite été envoyés à séquencer (société Genoscreen), puis alignés (Blastn avec les paramètres par défaut) sur le génome de référence *via* le site d'accès aux données du séquençage de la vigne du Génoscope pour vérifier la conformité du produit PCR amplifié.

## **2.2.6. Analyse du polymorphisme de séquence**

### **2.2.6.1. Détection du polymorphisme de séquence**

Les séquences Sanger obtenues par l'EPGV (fichier. ab1) ont été alignées entre elles et sur la séquence du gène ciblé chez le génome de référence PN40024 (séquence téléchargée sur le site d'accès aux données du séquençage de la vigne du Génoscope; pour vérifier l'adéquation entre la zone ciblée et la séquence obtenue en utilisant le logiciel Genalys version 2.8.3b pour Macintosh (Figure 2.7; Takahashi et al, 2003). Les séquences ont été nettoyées par rapport à leur qualité et par rapport aux alignements multiples: les INDEL ont été éditées à la main et les SNP ont été vérifiés et corrigés si nécessaire en prenant comme référence la séquence du génotype PN40024. Les séquences ont ensuite été formatées sous le format multifasta à l'aide d'un script perl maison, importées et stockées dans la base de données SNIPlay ouverte au public (A. Dereeper; <http://sniplay.cirad.fr>).

### **2.2.6.2. Analyse de la diversité nucléotidique, du déséquilibre de liaison et test de neutralité**

La diversité nucléotidique dans l'échantillon de génotypes SNPGrpMap de chaque fragment de gène a été analysée en utilisant le logiciel DNAsp version 5.10 (<http://www.ub.edu/dnasp/>;

Librado and Rozas, 2009). Ce logiciel permet à partir d'une séquence multifasta en code UPAC de reconstruire les haplotypes et de calculer le nombre de sites polymorphes: S, le nombre d'haplotypes: H, la diversité haplotypique: Hd, la diversité nucléotidique:  $\pi$  et de réaliser un test de neutralité: D de Tajima.

Au niveau de la région *Fleshless berry* (environ 1 Mb), la reconstruction des haplotypes pour la région entière à partir des données de séquences de 69 fragments séparés de 3 à 57 Kb n'a pas pu être réalisée à partir du logiciel DNAsp. Roger et Huff (2009) ont montré que le coefficient de corrélation génotypique calculé entre chaque SNP était un bon estimateur du coefficient de corrélation haplotypique habituellement utilisé pour estimer le déséquilibre de liaison. Afin de calculer le déséquilibre de liaison (LD) dans la région *Flb*, le coefficient de corrélation génotypique entre chaque SNP a été calculé à l'aide d'un script R développé par A. Siberchicot et C. Cierco (INRA, Toulouse) en utilisant les données génotypiques et donc sans reconstruction préalable des haplotypes. Le LD a été estimé dans la région entière par le carré du coefficient génotypique et sa P-value associée, et les résultats ont été visualisés grâce à un script Perl écrit par C. Guichard (URGV, Evry).

### 2.2.6.3. Choix des SNP

Les SNP et INDEL détectés par séquençage et utilisés pour génotyper un échantillon plus important de variétés ont été choisis en respectant le cahier des charges lié à l'utilisation du système SNPlex (Applied Biosystems):

- insertions/délétions (INDEL) <7 pb;
- autre polymorphisme (non ciblé) à une distance >15 bases du SNP ou INDEL ciblé;
- séquence de 70 pb, au minimum de 35 bp de part et d'autre de l'INDEL ou du SNP avec un minimum de polymorphismes dans les 70 pb flanquant l'INDEL ou le SNP ciblé;
- les séquences flanquant le SNP ou INDEL ont ensuite été alignées sur la séquence du génome afin de tester leur spécificité. Seules les séquences uniques ont été envoyées à Applied Biosystems.

Les filtres suivants ont également été appliqués dans le choix des SNP à cibler suivant les objectifs du génotypage:

- pour le génotypage des gènes candidats pour la variation de la taille de la baie, tous les SNP et INDEL avec une fréquence >5% dans le set des 26 vignes cultivées et présents sur des blocs haplotypiques différents ont été retenus;

- pour les 77 séquences de fragments de gènes répartis aléatoirement sur le génome (R. Bacilieri et J-P Péros, communication personnelle), 1 SNP avec une fréquence supérieure à 5% par gène dans l'échantillon des 26 vignes cultivées a été retenu;
- pour les gènes de la région *Fleshless berry*, 1 à 4 SNP ou INDEL par gène avec une fréquence >5% dans l'échantillon des 26 vignes cultivées et 7 vignes sauvages ont été retenus.

Après vérification des séquences et de la compatibilité de multiplexage des sondes par Applied Biosystems, quatre pannels SNplex ont été commandés: 1 de 47 SNP pour la région *Fleshless berry*, 1 de 48 SNP pour les gènes aléatoires répartis sur tout le génome et deux de 48 SNP chacun pour l'ensemble des gènes candidats pour la variation de la taille de la baie.

### 2.2.7. Génotypage

Les ADN destinés au génotypage ont été dilués afin d'obtenir la concentration la plus homogène possible autour de 5 ng dans 4 µl final par I. Le Clainche (Evry, URGV). Ils ont ensuite été répartis dans 4 plaques 96 contenant chacune deux puits d'H<sub>2</sub>O (témoins négatifs), deux puits avec de l'ADN de PN40024 (témoins positifs), et deux puits vides pour des contrôles internes. Les sondes SNplex, ainsi que les plaques d'ADN ont été transmises à l'EPGV (A. Bérard et A. Chauveau) qui ont réalisé les expériences de génotypage sur la plateforme.

Les résultats obtenus ont ensuite été analysés et édités si nécessaire à l'aide du logiciel Gene Mapper version 3.7 (Applied Biosystems) puis récupérés sous format Excel.

### 2.2.8. Analyses statistiques et tests d'association

#### 2.2.8.1. Tests d'association

Les tests d'associations ont été réalisés à l'aide du logiciel TASSEL version 2.1 (<http://www.maizogenetics.net/>). Deux modèles ont été utilisés pour rechercher des associations entre les marqueurs et les phénotypes d'intérêt:

- le modèle linéaire généralisé (*Generalized Linear Model* ou GLM) prenant en compte la structure génétique de chaque individu. Ce modèle s'écrit:  $P = G + Q + \varepsilon$  où P, G, Q et  $\varepsilon$  représentent respectivement la valeur phénotypique du caractère, le génotype du marqueur candidat, la structure et l'erreur résiduelle;

- le modèle linéaire mixte (*Mixed Linear Model* ou MLM) prenant en compte la structure génétique, ainsi que le lien de parenté entre les individus, ce modèle s'écrit:  
 $P = G + Q + K + \epsilon$  où K représente l'apparentement entre individus.

Le niveaux d'apparentement au sein de notre échantillon de génotypes a été calculée: pour la structure de l'échantillon (Q) grâce au logiciel Structure (Pritchard et al, 2000) à partir du génotypage de 20 marqueurs SSR répartis sur les 19 chromosomes (Le Cunff et al, 2008; R. Bacilieri résultats non publiés) et pour la matrice d'apparentement des individus deux à deux (K) grâce au logiciel TASSEL version 2.1 à partir de 45 SNP répartis aléatoirement le long du génome (2 à 3 SNP par chromosome) provenant d'un des quatre panels SNPLex. Le seuil de significativité a été choisi à 0.05, après la prise en compte des tests multiples par la correction de Bonferroni.

#### **2.2.8.2. Analyses statistiques**

Les analyses statistiques des données phénotypiques ont été réalisées en utilisant le programme R version 2.10.1 (<http://www.r-project.org/>). Des régressions linéaires ainsi que la valeur du coefficient de corrélation de Spearman au carré ( $R^2$ ), ont été calculées pour toutes les variables deux à deux.

### **3. Analyse phénotypique fine de la variation de la taille de la baie chez *V.* *vinifera***



*Photographie d'une grappe de raisins quelques semaines après nouaison.*





### 3.1. Introduction

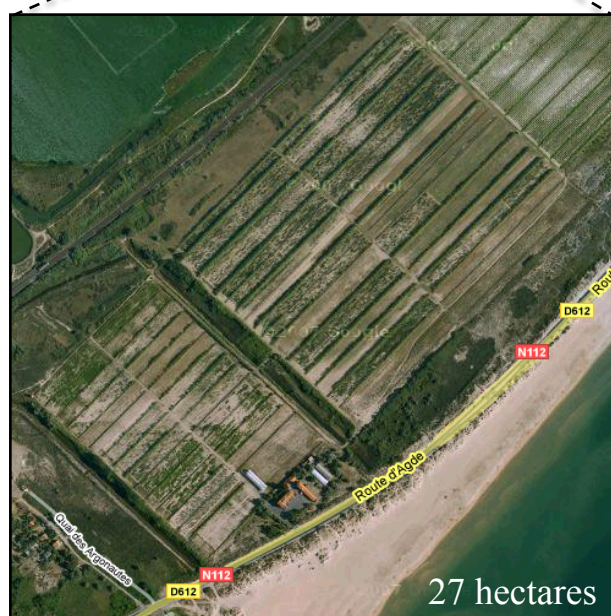
Actuellement, parmi les espèces à fruits charnus, seule la tomate a été soumise à des recherches approfondies concernant la variabilité de la taille du fruit. Ces études ont montré que les trois paramètres suivants: le nombre de cellules de la baie, leur volume et le nombre de carpelles expliquent la variation de la taille du fruit entre le compartiment sauvage et cultivé et également la variation à l'intérieur du compartiment cultivé (Frary et al, 2002; Cheniclet et al, 2005; Cong et al, 2008).

Chez la vigne cultivée (*V. vinifera* ssp *sativa*), il existe également une forte variabilité de la taille des fruits (Boursiquot et al, 1995). Cependant, cette variabilité reste très peu documentée au sein du compartiment sauvage et, en ce qui concerne le compartiment cultivé, l'étude des mécanismes sous-jacents se limite souvent à des descriptions phénotypiques concernant des différences entre clones chez quelques variétés (Fernandez et al, 2006a et 2006b). A l'heure actuelle, aucune description fine des baies sur un échantillon représentatif de la vigne cultivée n'a été réalisée.

L'objectif de cette partie de ma thèse, a été de décomposer le caractère taille de la baie en sous-caractères tels que le nombre et le volume des cellules de la pulpe, le nombre de carpelles, mais également d'analyser la contribution du nombre et du volume des cellules présentes initialement dans l'ovaire juste avant l'anthèse. Une telle description permettrait de mieux comprendre les différents mécanismes à l'origine de la variation de la taille du fruit chez la vigne cultivée ainsi que le(s) étape(s) du développement où se détermine la taille finale de la baie.

Pour cela, je me suis basée sur une collection de 278 génotypes non apparentés et représentant 80% de la variabilité génotypique du compartiment cultivé de *V. vinifera* estimée à l'aide de 20 marqueurs SSR (S. Nicolas, T. Lacombe et L. Le Cunff communication personnelle). Cette collection a également été constituée en tenant compte de sa structuration en 3 sous-populations (Le Cunff et al, 2008). Nous avons ajouté 25 autres génotypes cultivés permettant de compléter la diversité génotypique et phénotypique de l'échantillon. Enfin, j'ai également étudié l'ensemble des caractères chez un génotype sauvage (Lambrusque de Grésigny 1), afin d'avoir une première idée des différences au niveau de l'ovaire et de la baie entre la vigne sauvage et cultivée.

Dans ce chapitre, je présente les résultats que j'ai obtenus sur le phénotypage de ces caractères.



**Figure 3.1:** Localisation des différents sites dans lesquels les ovaires et les baies ont été prélevés et préparés pour l'ensemble des mesures réalisées.

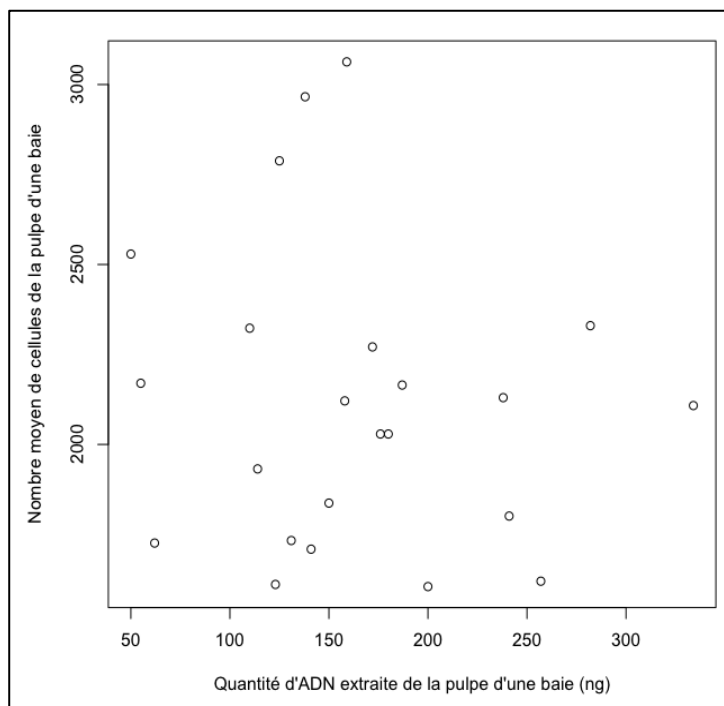
Les ovaires et baies ont été collectés respectivement de mai à juin et de juin à septembre, en fonction des dates de floraison, véraison et maturation des génotypes, dans l'unité expérimentale du Domaine de Vassal à Marseillan, rattachée à l'INRA de Montpellier (Figure 3.1). Les mesures ont ensuite été réalisées à l'INRA de Montpellier au sein de l'équipe DAVEM (Diversité et Adaptation de la Vigne et des Espèces Méditerranéennes) dans l'UMR AGAP dirigée par le Dr Patrice This et dans la Plateforme d'Histo-cytologie et d'Imagerie cellulaire Végétale (PHIV) du CIRAD dirigée par le Dr Jean-Luc Verdeilau (Figure 3.1).

Ma première préoccupation a été d'essayer de trouver un estimateur du nombre et du volume des cellules de la pulpe des baies qui puisse être analysé sur un grand nombre de génotypes dans le cadre du stage de Master 1 de Damien Frank que j'ai co-encadré.

### **3.2. Définition d'une stratégie pour l'estimation du nombre de cellules de la baie**

Le comptage cellulaire par une méthode classique (coupes histologiques) n'est pas adapté au criblage d'une collection de 304 individus, surtout si on analyse au minimum 3 ovaires et 3 baies par génotype pour avoir des valeurs représentatives de ce dernier. D'après les travaux d'Ojeda et al (1999), la quantité finale de cellules de la baie est fixée chez la vigne quelques semaines avant véraison. La croissance des baies pendant la maturation, stade suivant la véraison, est en effet uniquement due à des phénomènes de grandissement cellulaire sans phénomène d'endoréduplication comme chez la tomate à l'origine de l'augmentation du volume cellulaire. La quantification de l'ADN présent dans la baie juste avant véraison, détermine donc indirectement le nombre de cellules de la baie (Ojeda et al, 1999). Cette méthode a déjà été utilisée dans plusieurs études comme estimateur du nombre de cellules de la pulpe chez la vigne (Ojeda et al, 1999; Fernandez et al, 2006a). Les cellules de la pulpe représentant l'immense majorité des cellules de la baie, les cellules de la pellicule étant négligeables, l'ADN extrait à partir de baies épépinées correspond à l'ADN de la pulpe. La quantification de l'ADN présent dans la pulpe de chaque génotype serait une méthode simple et rapide, permettant d'étudier la variabilité du nombre des cellules dans notre collection de 304 génotypes et c'est ce que nous avons voulu tester.

Dans un premier temps, la quantification de l'ADN présent dans la pulpe a été réalisée



**Figure 3.2:** Corrélation entre le nombre moyen de cellules comptées sur des coupes transversales de baies observées au microscope optique et la quantité d'ADN extraite par baie dans un échantillon de 26 génotypes cultivés.

sur trois fois 2 g de poudre de baies provenant de 10 baies dures récoltées sur une grappe contenant au moins une baie véree, sur un échantillon de la collection, contrastée pour la taille de la baie, comprenant 26 génotypes. La quantité d'ADN a ensuite été ramenée au poids de la baie. Des coupes de trois baies par génotype récoltées au même stade ont également été réalisées pour ces mêmes génotypes, afin de vérifier si la quantification de l'ADN est un bon estimateur du nombre de cellules de la pulpe.

Nous n'avons pas pu mettre en évidence de corrélation entre le nombre de cellules de la pulpe et la quantité d'ADN extraite (Figure 3.2). La quantification de l'ADN des baies au stade pré-véraison ne semble pas être un estimateur fiable du nombre de cellules dans notre échantillon. Ce résultat peut s'expliquer par le fait que dans les études précédentes (Ojeda et al, 1999 et Fernandez et al, 2006a), la quantification de l'ADN avait permis de comparer le nombre de cellules à différents stades du développement d'un même génotype. Ici, nous comparons la quantité d'ADN entre génotypes. Malgré une bonne répétabilité des extractions pour un même génotype, une différence de composition de la pulpe est certainement à l'origine d'une différence de rendement d'extraction d'ADN entre les génotypes, rendant impossible une comparaison de la quantité d'ADN entre eux, ce qui pose problème pour le phénotypage de ce caractère dans la collection entière. Il faut noter d'ailleurs que le broyage des baies a nécessité l'usage d'un broyeur particulier (voir matériel et méthodes) qui ne peut broyer qu'un génotype à la fois, soit l'extraction de l'ADN d'une dizaine d'échantillons au maximum par jour.

Nous avons donc décidé d'analyser le nombre et le volume des cellules des baies et des ovaires à partir de coupes histologiques chez les 26 génotypes représentatifs de la variabilité de la taille de la baie chez la vigne cultivée, ainsi que le génotype sauvage, et d'extrapoler les résultats à l'ensemble de la collection. Cependant, toujours dans un souci de gain de temps à terme, plusieurs techniques d'histologie demandant un temps de préparation des échantillons plus ou moins long ont été testées:

- coupe au vibratome sur du matériel frais et fixé ;
- coupe au cryo-microtome sur du matériel fixé et congelé ;
- coupe au microtome sur du matériel fixé et inclus en résine.

Malheureusement, les deux premières techniques demandant peu de préparation des échantillons (3 à 4 jours) se sont avérées inadaptées à la baie de vigne. En effet, les cellules de la pulpe ont des parois et membranes cellulaires très fines même au stade pré-véraison, qui se rompent très facilement sur du matériel frais et ne supportent pas la congélation/décongélation.

La dernière technique qui demande plus d'un mois entre la récolte du matériel végétal et l'acquisition des images, a donc finalement été utilisée pour estimer le nombre et la taille des cellules de l'ovaire et de la pulpe de la baie.

### **3.3. Description de la variabilité de la taille de la baie chez *V. vinifera***

L'analyse des données macroscopiques de phénotypage obtenues sur la collection de 305 génotypes et microscopiques obtenues sur l'échantillon de 26 génotypes est présentée sous la forme d'un article qui a été soumis le 18 mars 2011 dans *Journal of Experimental Botany*. Les figures et tableaux supplémentaires de l'article sont regroupés dans l'annexe 1.

## Genetic variability of the berry size in grapevine (*Vitis vinifera* L.) involves ovary and early berry development associated factors

Cléa Houel<sup>1\*</sup>, Marie-Laure Martin-Magniette<sup>1,2</sup>, Stéphane D. Nicolas<sup>3,4</sup>, Thierry Lacombe<sup>3,5</sup>, Loïc Le Cunff<sup>3,6</sup>, Damien Franck<sup>1</sup>, Laurent Torregrosa<sup>3</sup>, Geneviève Conéjéro<sup>7</sup>, Sandrine Lalet<sup>5</sup>, Patrice This<sup>3</sup> and Anne-Françoise Adam-Blondon<sup>1\*</sup>

<sup>1</sup> Unité Mixte de Recherche en Génomique Végétale, INRA UEVE ERL CNRS, 2 rue Gaston Crémieux, BP 5708, 91057 Evry, France

<sup>2</sup> Unité Mixte de Recherche en Mathématiques et Informatique Appliquées, INRA AgroParisTech, 75231 Paris, France

<sup>3</sup> TGU AGAP – Equipe DAVEM, INRA-Montpellier SupAgro, 2 place Pierre Viala, 34060 Montpellier, France

<sup>4</sup> Unité Mixte de Recherches en Génétique Végétale, INRA CNRS Univ Paris-Sud AgroParisTech, Ferme du Moulon, 91190 Gif sur Yvette, France

<sup>5</sup> INRA – Domaine de Vassal, route de Sète, 34340 Marseillan-Plage, France

<sup>6</sup> Unité Mixte Technologique Géno-Vigne, INRA IFV SupAgro, Domaine de l'Espiguette, 30240 Le Grau du Roi, France

<sup>7</sup> UMR Biochimie et Physiologie Moléculaire des Plantes, Histology and Plant Cell Imaging Platform, CIRAD, avenue Agropolis, 34398 Montpellier, France

### Abstract

Like in tomato, berry weight is variable in grapevine, ranging from less than 1 g to 10 g. In tomato, berry weight depends on cell number and volume but also on carpel number. The aim of this work was to decipher the main factors underlying berry size variation in a highly diverse collection of cultivated grapevines. Cell divisions before and after anthesis and cell expansion after anthesis were the major determinants of flesh weight variation between varieties. Carpel number varied between varieties, with two and three carpels per ovaries. However, this trait and also the seed weight did not clearly seem to contribute to berry size variation although a positive correlation was found between the seed weight and number and the berry weight at the intra-genotypic level, in agreement with previous results. The factor explaining the transition between the wild and the cultivated genotypes regarding to berry size is unclear from the preliminary data presented here, as none of the measured traits seemed clearly out of the range of variation of the cultivated genotypes carrying small berries. This study highlights the common (cell number and volume) and specific mechanisms (carpel number) underlying the fruit size variation in fleshy fruits and their extent of contribution, depending on the fruit species.

## Background

Domestication of most crops has correlated with a dramatic increase in size of the harvested organs (Cong et al., 2008) that led to distinguish domesticated taxa from their wild ancestors (Purugganan and Fuller, 2009). Moreover, selection after domestication resulted in a large phenotypic diversity in varieties that can exceed the range of variation in their wild ancestors (Purugganan and Fuller, 2009). For instance, cultivated tomato *Solanum lycopersicum* plants can produce fruits as much as 1,000 times heavier than the wild tomato *Solanum esculentum* (Tankley, 2004). A similar situation is observed between the wild grapevines (*Vitis vinifera* subspecies *sativa*) that produce berries weighting less than 1 g and some cultivated grapevines (*Vitis vinifera* subspecies *sylvestris*) that produce berries that can weigh as much as 10 g (Boursiquot et al., 1995).

In fleshy fruits like tomato or grapevine, the berries develop from ovary after fertilization. The ovary wall becomes the pericarp, which is composed of three distinct cell layers: the epicarp that corresponds to the skin, the mesocarp that corresponds to the flesh, and the endocarp that corresponds to the cell layers in contact with the seeds and that can be either fleshy in berries or lignified in stone fruits (Coombe, 1976; Ollat et al., 2002). Like in most other fleshy fruits, the growth of the grape berries follows a double sigmoid curve, corresponding to three developmental stages (Coombe, 1976). After flowering, during stage I, the berry growth is due to cell mitotic divisions and cell expansion. Approximately four to six weeks post-anthesis, cell division ceases and only cell expansion subsists. At the end of stage I, all the tissues composing the seeds are formed. The stage II corresponds to a slow growth phase. The end of the stage II, known as *véraison*, is easily recognizable: the texture of the berry changes from hard to soft and the skin changes of color easily recognizable for the red varieties. Finally, the berry growth restarts in stage III only through cell enlargement (Coombe, 1976; Ojeda et al., 1999).

In tomato, less than ten loci seem to be responsible of the majority of changes in tomato fruit size (Tankley, 2004) and this variation is first due to the variation of cell volume and number, and second to the carpel number variation (Tankley, 2004; Cheniclet et al., 2005; Cong et al., 2008). A Quantitative Trait Locus (QTL) was shown to control the variation of the ovarian cell number before anthesis and to explain 30% of the difference in fruit size between large domesticated tomatoes and their small-fruited wild ancestor (Frary et al., 2000; Cong et al., 2002). The increase in carpel number in the cultivated tomato varieties has been shown to cause as much as 50% increases in fruit size (Cong et al., 2008). Finally, to date two genes



involved in fruit weight variation were cloned in tomato: *FW2.2* which is a negative regulator of cell division in ovary (Cong et al., 2002) and *FASCIATED* which controlling the number of carpels in flower (Cong et al., 2008). In grapevine, few genetic studies of the determinism underlying berry size variation were carried out and generally in association with the seedless trait that shows strong interactions with the berry size and made difficult the detection of stable QTL unlinked with the seedless locus (Doligez et al., 2002; Fanizza et al., 2005; Cabezas et al., 2006; Mejia et al., 2007; Mejia et al., 2011). In addition, Fernandez et al., (2006a) showed that cell enlargement might explain the differences of berry size between three cultivated varieties and their clonal variants differing for berry size. In one of these clones, showing the fleshless berry extreme phenotype, the number of cells was also affected (Fernandez et al., 2006a and 2006b). In grapevine, no detailed comparative description of the berry development in small versus large berry genotypes has been carried out so far in a large sample of genotypes like in tomato (Cheniclet et al., 2005).

The aim of this work was to decipher the main anatomical factors underlying berry size variation in *Vitis vinifera* (*V. vinifera*) through the study of ovarian and berry components at different stages of development in a highly diverse collection of genotypes. This will allow us (i) to better compare factors of berry size variation to the factors of tomato size variation, which will be useful for transferring knowledge from the model species for fleshy fruit to a perennial species and (ii) to propose the best and easiest component(s) to be measured for fine phenotyping for berry size in future genetic studies.

## Materials and Methods

### Plant material

The samples of berries and ovaries were harvested from 304 *V. vinifera* ssp. *sativa* genotypes listed in Table S1 and maintained in the French National Grapevine Germplasm Collection (INRA Domaine de Vassal, Montpellier, France; <http://www1.montpellier.inra.fr/vassal/>). Based on the genotyping of this collection using 20 Short Sequence Repeat (SSR) (Lauco et al., 2011), 278 genotypes were chosen to minimize the familial relatedness within three sub-populations obtained after analysis of the genotyping results with the Structure software (Pritchard et al., 2000). These 278 genotypes contain 93 wine cultivars originating from Eastern Europe (WE), 92 wine cultivars originating from the Western Europe (WW) and 93 table cultivars originating from Eastern Europe (TE), and represent 80% of the SSR genetic diversity of the French National Grapevine Germplasm Collection. Seventeen genotypes



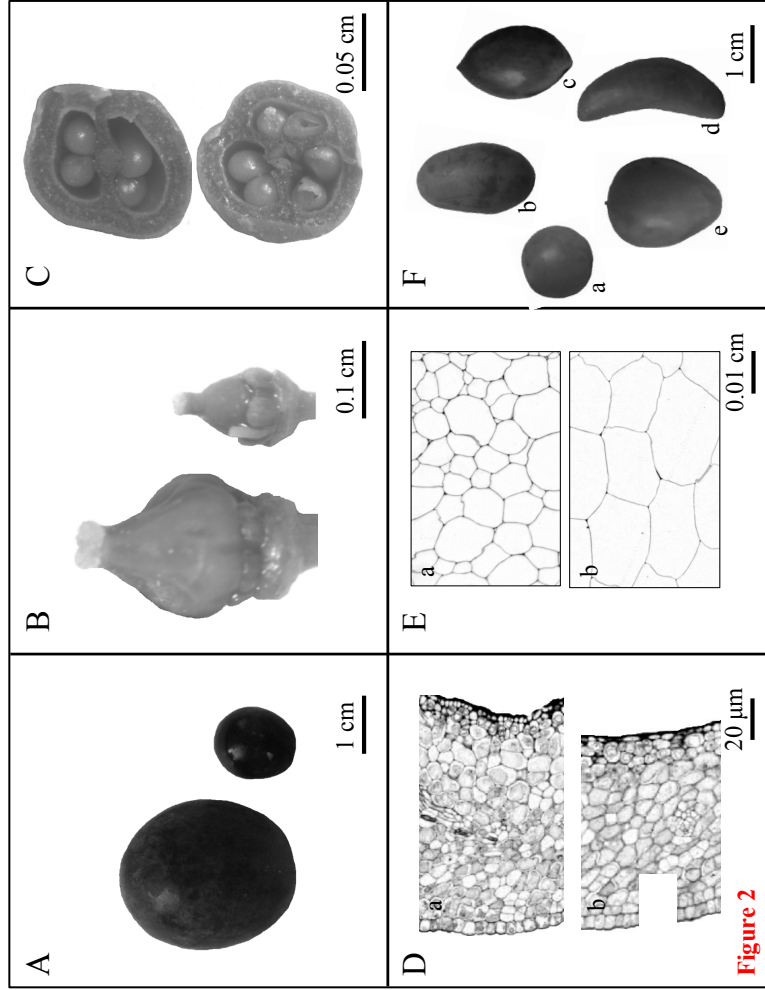
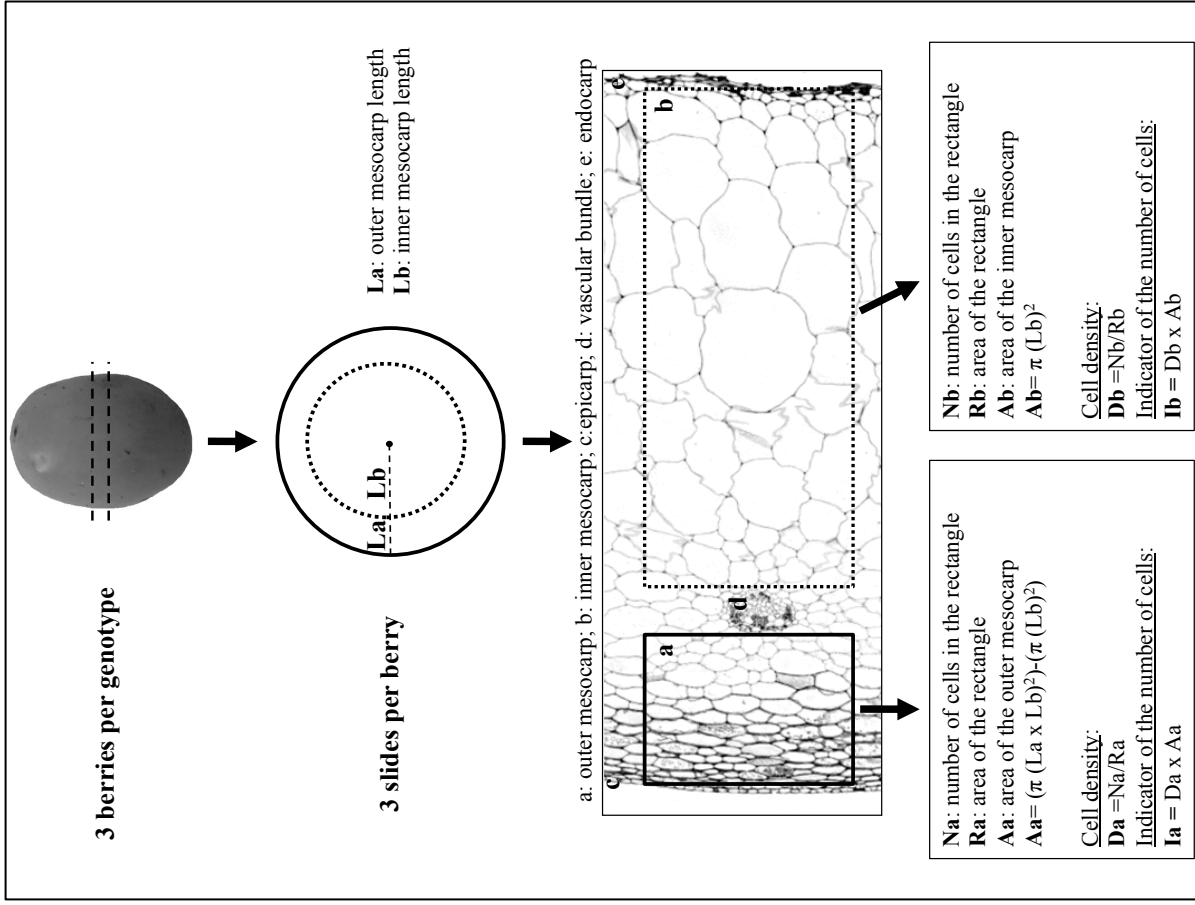
maximizing the genetic diversity of cultivated *V. vinifera* sampled by Le Cunff et al. (2008) and described by Houel et al. (2010) were added to complete the sample. Eight other cultivated *V. vinifera* were added in order to improve the sampling of the phenotypic diversity for berry size in cultivated grapevine. Finally, one wild accession (*V. vinifera* ssp. *sylvestris*: Lambrusque de Grésigny 1) was studied as a wild out group.

### **Berry and ovary phenotyping**

For each genotype, the average berry weight (on 30 berries) and diameter (on ten berries) were measured from hard berries at *véraison* cut at the pedicel base (stage 81 from BBCH international scale; Lorenz et al., 1994) and from mature berries cut at the pedicel base (40 days after *véraison*; stage 89 from BBCH international scale). At the *véraison* stage, the berries have reached their final number of cells. The berries were randomly sampled within the clusters on at least two different clusters from two plants in 2008 (231 genotypes) and in 2009 for the genotypes which had not berry the year before (23 genotypes and 18 genotypes already phenotyped in 2008 as controls). The length (L) and the width (W) of the berries were measured from photographs using the ImageJ® version 1.43 software for Mac OS X (<http://rsbweb.nih.gov/ij/download.html>). The berry volume was estimated as an ellipsoid volume using the formula:  $V = \frac{4}{3} \cdot \pi \cdot L \cdot W^2$  and the berry shape was approached as the ratio between the length and the width of the berry. The number of seeds and the seed weight per berry were measured from ten berries at *véraison* (at this stage the seeds are entirely formed (Ojeda et al., 1999) and easier to separate from the flesh than at maturity). The flesh weight at maturity and *véraison* stages were estimated by subtracting the average seed weight to the average berry weight.

For a representative sample of 26 genotypes from the three sub-populations (without seedless varieties; Table S1), the berry weight, the seed weight and number were also measured individually from ten berries at maturity, to study the intra-genotypic correlation between traits. The length and width of ovaries were measured from ten closed flowers (entirely formed but not fertilized) per genotype, sampled on at least two different inflorescences from two plants showing a few open flowers (stage 61-62 from BBCH international scale). The ovary volume and shape was approached using the same method than for the berries. The ovaries were cut widthwise in two pieces in order to count their number of carpels under a binocular.

### **Berry and ovary preparation for microscopy**



**Figure 2.** Variation of ovarian and berry traits in cultivated grapevine. (A) The largest and the smallest berries at maturity respectively *cv.* Barlinka (left) and *cv.* Domina (right). (B) The largest and the smallest ovaries at anthesis respectively *cv.* Kozouzioum (left) and *cv.* Ramisco (right). (C) The range of carpel number variation in flowers from cultivated grapevine. (D) Transversal sections of ovaries at anthesis from *cv.* Dabouki from Group D with large berries and (b) from *cv.* Fer from Group A with small berries, highlighting the difference in cell number. (E) Transversal sections of berries at véraison showing the difference in cell volume between (a) *cv.* Fer and (b) *cv.* Dabouki. (F) Examples of fruit shape variation at véraison: (a) *cv.* Moscato di Terracina from Group B, (b) *cv.* Aïn Kelba from Group B, (c) *cv.* Teta de Vaca from Group C, (d) *cv.* Coarna Alba from Group B, and (e) *cv.* Emerald from Group C.

**Figure 1.** Method for the calculation of the indicators of the number and the volume of cells in the mesocarp of berries at véraison. Ia and Ib are respectively the indicators of the number of cells in the outer and inner mesocarps. Da and Db are respectively the indicators of the cell volume in the outer and inner mesocarps.

For the representative sample of 26 genotypes from the three sub-populations (Table S1), three closed flowers and three berries at *véraison* per genotype were randomly chosen among the previous 30 collected berries, vacuum infiltrated three times for 10 minutes (min) in a cold solution of fixation (sodium phosphate buffer 0.1 M pH 7.2, paraformaldehyde 2%, glutaraldehyde 1% and 1% caffeine) and maintained in a fresh fixative solution for at least 48 hours (h) at 4 °C. For each berry, the pieces with the larger diameter of the berry (Figure 1) were further prepared. They were then washed in a solution of ethanol 70% before removing the seeds and dehydrated through a series of incubations in ethanol: ethanol 70% for 1 h, ethanol 95% for 1 h and finally, ethanol 100% for 24 h at 4 °C. The dehydrated samples were incubated in Technovit 7100 cold-curing resin for soft tissues (Kulzer, Labonord, France) for at least two weeks shaking at 4°C. After addition of 1% methyl methacrylate (hardener solution) the resin was allowed to polymerize at 37°C for 24 h in a dry atmosphere. After inclusion, the samples were cut into 3 µm sections using a LEICA<sup>TM</sup> RM2255 microtome, stained in a solution of Alcian blue 8 GX 0.1% and sodium citrate 0.01 M / hydrogen chloride (HCl) 0.01 M buffer pH 3.5 for 24 h in order to allow visualizing the cell membranes, washed twice in sodium citrate / HCl buffer pH 3.5 and finally mounted on slides with Isomount medium (Labonord, France).

### **Histological studies of berries and ovaries**

The berry and ovary sections were observed under a microscope (DM 4500, Leica<sup>TM</sup>). Three objectives were used: ×5 for measuring berry outer and inner mesocarp length and for counting the berry cell number, ×10 for measuring ovary outer and inner mesocarp length and ×20 for counting the number of ovarian cells. Images were taken on each slide using a camera (Q-Imaging; Micropublisher version 3.3 RTV) coupled to the microscope and driven by the Volocity® version 5.4.1 software for Mac OS (<http://www.cellularimaging.com/products/voloccity/>). The three best slices for each three ovaries and berries per genotype were chosen (see Figure 1 for berries), and if necessary the berry slices were reconstructed from images by photomerge using the software Photoshop® version CS3. The outer and inner mesocarp cell densities were estimated as the number of cells present in a defined area chosen in order to capture the maximum of cell size variability of the corresponding mesocarp at the largest diameter of the berry (Figure 1), using the software ImageJ®. The cell density was used as an indicator of the cell volume because it is inversely correlated with the cell volume. The outer and inner mesocarp cell numbers were

**Table 1.** Average, minimum and maximum values for the parameters measured on ovaries before anthesis and berries at *véraison* and maturity stages on cultivated *V. vinifera* genotypes.

	Number of genotypes	Average	Standard deviation	Minimum value	Maximum value	Variation range*
<b>OVARIES</b>						
Carpel number	180	2.1	0.2	2.0	2.8	1.4
Length (cm)	141	0.13	0.02	0.09	0.19	2
Width (cm)	141	0.10	0.02	0.06	0.15	2.5
Volume (cm <sup>3</sup> )	141	0.0008	0.0003	0.0002	0.0022	11
Shape estimator	141	1.3	0.1	1.0	1.6	1.6
Outer mesocarp cell density (cell.µm <sup>-2</sup> )	26	0.0097	0.0022	0.0050	0.0133	2.6
Inner mesocarp cell density (cell.µm <sup>-2</sup> )	26	0.0092	0.0014	0.0064	0.0121	1.9
Total mesocarp cell density (cell.µm <sup>-2</sup> )	26	0.0090	0.0017	0.0053	0.0123	2.3
Indicator of cell number	26	1608	499	677	2528	3.7
<b>BERRIES</b>						
Seed number	254	1.9	0.7	0.0	3.6	3.6
Seed weight (g)	254	0.10	0.04	0.00	0.22	22
<i>Véraison</i>						
Flesh weight (g)	254	1.22	0.62	0.39	3.76	10
Length (cm)	254	1.4	0.3	1.0	2.6	2.6
Width (cm)	254	1.3	0.2	1.0	1.9	1.9
Volume (cm <sup>3</sup> )	254	2.8	1.4	0.5	4.7	10
Shape estimator	254	1.1	0.2	1.0	1.9	1.9
Outer mesocarp cell density (cell.µm <sup>-2</sup> )	25	0.00060	0.00022	0.00024	0.00117	4.9
Inner mesocarp cell density (cell.µm <sup>-2</sup> )	25	0.00004	0.00001	0.00002	0.00010	5
Total mesocarp cell density (cell.µm <sup>-2</sup> )	25	0.00032	0.00011	0.00013	0.00060	4.6
Indicator of cell number	25	2101	413	1605	3063	1.9
<i>Maturity</i>						
Flesh weight (g)	247	2.77	1.41	0.98	10.14	10
Length (cm)	247	1.8	0.4	1.0	3.3	3.3
Width (cm)	247	1.7	0.3	1.0	2.7	2.7
Volume (cm <sup>3</sup> )	247	1.4	0.6	0.5	11.5	23
Shape estimator	247	1.1	0.1	1.0	2.1	2.1

\* Ratio between the maximum value and the minimum value

estimated using the cell density multiplied by the area of the corresponding mesocarp, and represented the number of cell on a slice of ovary or berry (Figure 1).

### Statistical analysis

All variables were log-transformed in order to better fit a Gaussian distribution. Statistical analyses were performed using the R software version 2.10.1 (<http://www.r-project.org/>). Linear regression and Pearson correlation value (R) were estimated between each pair of variable. A Bonferroni correction was applied to control the Family Wise Error Rate at level 0.05.

### Ward method clustering

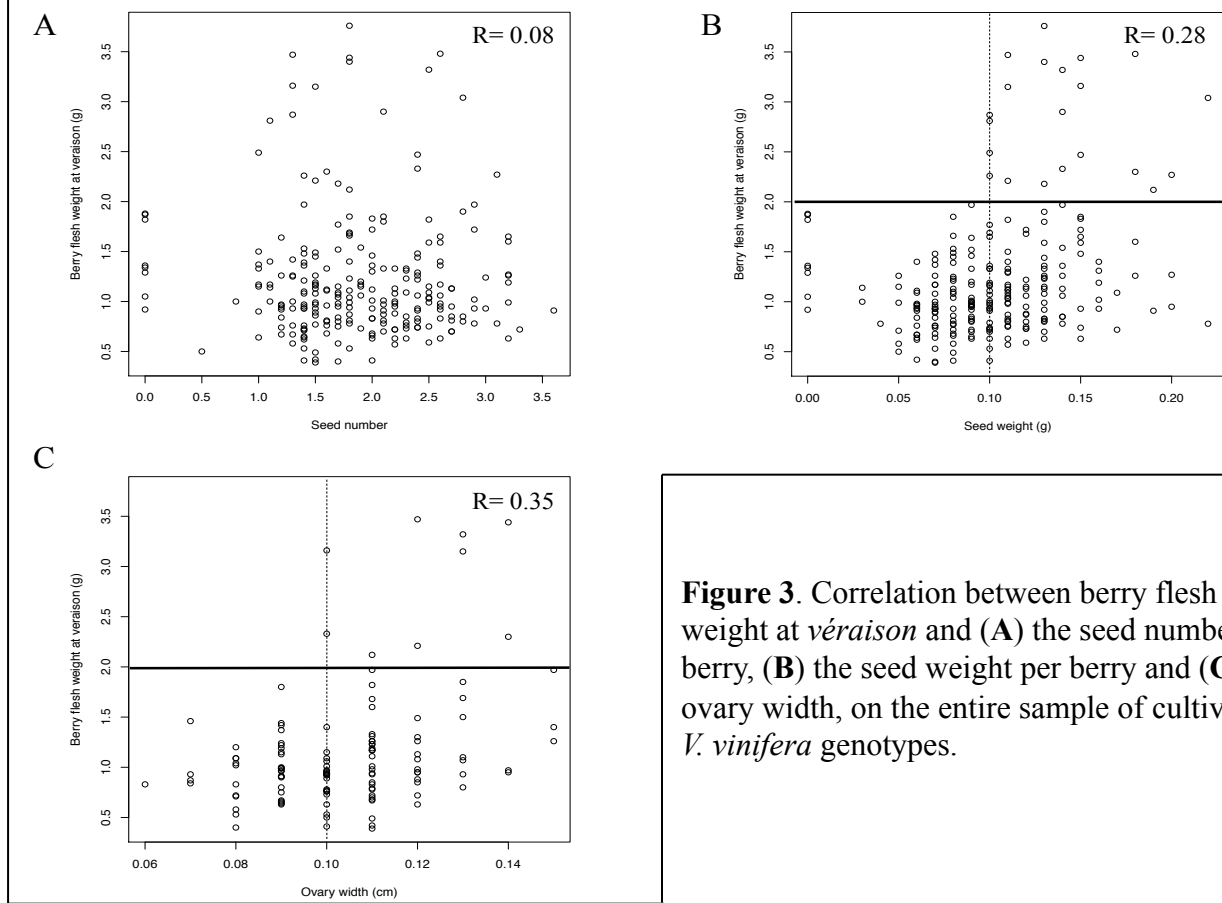
Data are clustered using the Ward method. It is an agglomerative clustering method. The similarity distance is the euclidian distance and the clustering criterion is based on the error sum of squares (E), which is defined as the sum of the squared distances of individuals from the centre of gravity of the cluster to which they have been assigned. Initially, E is zero, since every individual is in a cluster of its own. At each stage the link created is the one that makes the least increase to E.

## Results

### Distribution of the traits in the whole sample of cultivated genotypes and analysis in relation with berry size

Two hundred fifty four out of the 303 cultivated genotypes (Table S1) could be phenotyped for ovarian and/or berry traits during the two years. The missing data for 17% of all the genotypes were mainly due to *millerandage* (short berries due to incomplete pollination and development of the grape cluster), rabbits or bird invasions and in a few cases to too young plants not producing enough flowers and berries. All the parameters that were measured on ovaries just before anthesis, on berries at *véraison* (end of the stage where cell divisions and a few cell extension occur; Ojeda et al., 1999) and at maturity (stage where only cell extension was observed by Ojeda et al., 1999) are summarized in Table 1. Labor intensive histological parameters were only measured on a representative sample of 26 genotypes from the three sub-populations while the other parameters were measured on the panel of 254 cultivated grapevines (Table S1).

The berry flesh weight as well as the berry volume at maturity varied respectively with a ten and 23 factor (Table 1), from 0.98 g and 0.5 cm<sup>3</sup> (cultivar (*cv.*) Domina; Figure 2A) to 10.14



**Figure 3.** Correlation between berry flesh weight at *véraison* and (A) the seed number per berry, (B) the seed weight per berry and (C) the ovary width, on the entire sample of cultivated *V. vinifera* genotypes.

**Table 2.** Intra-genotypic correlation between berry flesh weight at maturity and the seed number or the seed weight per berry on the sample of 26 cultivated genotypes.

Variety name	Accession number	Pearson correlation	
		Seed number / berry flesh weight	Seed weight / berry flesh weight
Baresana = Korithi aspro	1247Mtp1	0.71	0.85*
Primitivo	1277Mtp6	0.96*	0.97*
Ohanes	1365Mtp1	0.85*	0.84*
Alvarelhão	1481Mtp1	0.80*	0.83*
Carcajolo	1531Mtp1	0.73*	0.67*
Berbecel	1573Mtp1	0.70	0.76*
Kövidinka	1578Mtp1	0.85*	0.91*
Mondeuse	161Mtp1	0.85*	0.75*
Chaouch blanc	1673Mtp1	0.37	0.75*
Verdeca = Lagorthi	2287Mtp1	0.85*	0.87*
Roussaitis	2312Mtp1	0.70	0.78*
Alexandroouli	2500Mtp1	0.72*	0.79*
Savagnin blanc	257Mtp1	0.56	0.71
Fer	353Mtp1	0.77*	0.87*
Courbu	413Mtp1	0.87*	0.89*
Razachie rosie	433Mtp1	0.83*	0.93*
Mauzac	443Mtp1	0.80*	0.61
Chasselas	585Mtp1	0.31	0.40
Olivette Blanche	667Mtp1	0.41	0.61
Nehelescol	727Mtp1	0.86*	0.84*
Dabouki <sup>s</sup>	735Mtp1	—	—
Heptakilo	743Mtp1	0.79*	0.87*
Coarna alba	749Mtp1	0.77*	0.79*
Ugni blanc	74Mtp1	0.63	0.92*
Calitor	77Mtp1	0.40	0.79*
Ahmar Derani	848Mtp1	0.72*	0.80*

\* Significant correlation with a Bonferroni correction at 0.05 (remark: in all the cases where no correlation was found, there was only two classes of seed weight or seed number)

<sup>s</sup> Missing data due to genotype not producing enough berries

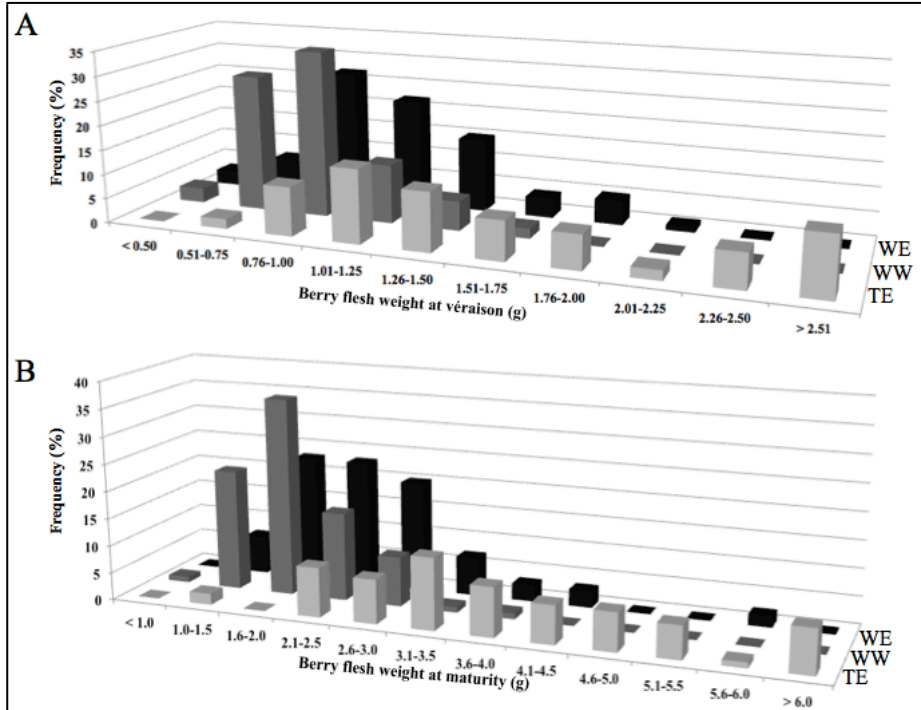


g and 11.5 cm<sup>3</sup> (for cv. Barlinka; Figure 2A). The berry flesh weight at *véraison* varied exactly in the same range than at maturity whereas the berry volume variation range was a twice smaller at *véraison* than at maturity (Table 1). The volume of ovary just before anthesis also showed wide differences between the genotypes, with a 11-fold between the smallest (0.002 cm<sup>3</sup> for cv. Ramisco; Figure 2B) and the largest (0.022 cm<sup>3</sup> for cv. Kozouzioum; Figure 2-B; Table 1). The number of carpels per ovary varied from two to three within and between the genotypes (Table 1, Figure 2C), with 29% of the genotypes showing flowers with three carpels. As already mentioned by Ollat et al., (2002), the average seed number per berry in seeded varieties was 1.9 (Table 1) with a nearly four-fold variation among the genotypes. The seed weight showed also a larger range of variation from 0.03 g to 0.22 g in seeded varieties.

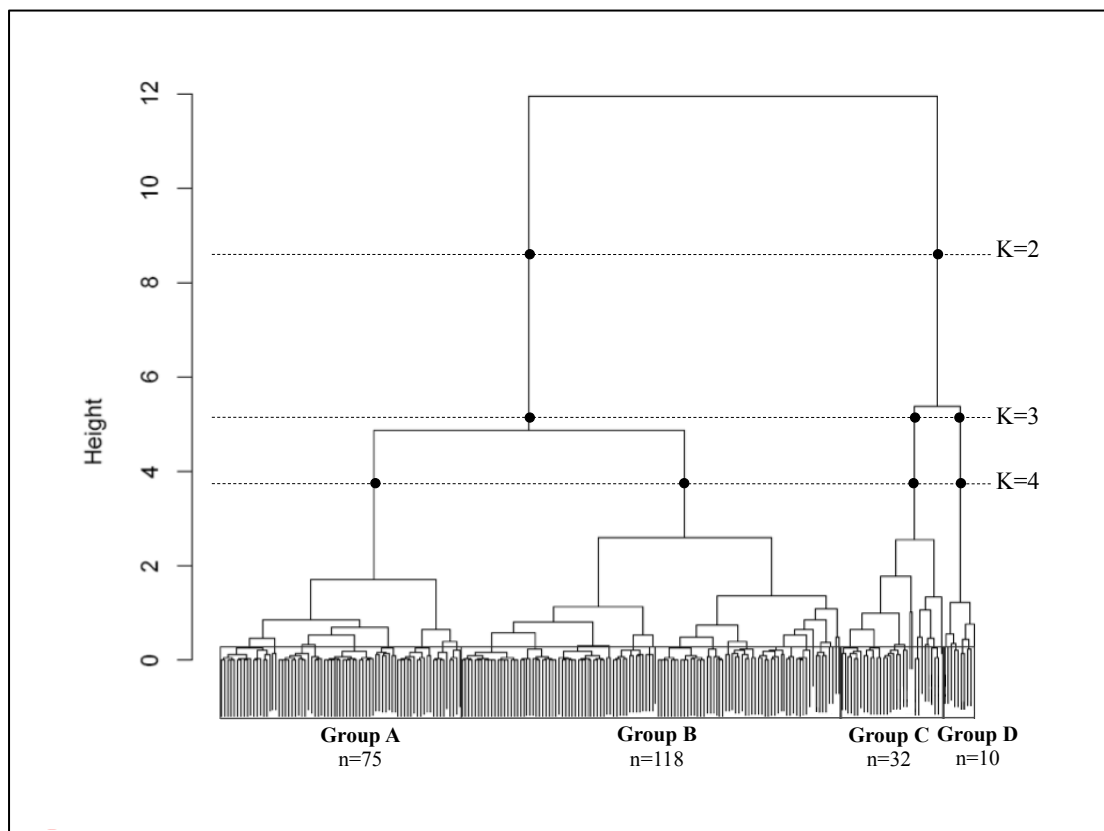
A strong correlation ( $R= 0.86$ ) was observed between the berry flesh weight at *véraison* and maturity. This showed that most of the events accounting for berry size variation between genotypes were determined before *véraison*. Moreover, at maturity the berry flesh weight and the berry volume were highly correlated traits ( $R= 0.91$ ). Finally, there were more missing data at the maturity stage due to pathogen attacks or birds feeding than at earlier stages of development. Therefore, the results obtained at *véraison* were preferentially used in the analysis and when the results at maturity were to be considered, only the berry flesh weight was analyzed. Finally, there was a strong correlation ( $R= 0.94$ ) between the average berry flesh weight at *véraison* in 2008 and 2009 from the 18 genotypes used as control (Figure S1). This result allowed us to merge the 2008 and 2009 data measured at *véraison*.

There was not significant inter-genotypic correlation between the seed number or the seed weight per berry and the berry flesh weight at *véraison* (Figure 3A and B) or maturity. Indeed, genotypes with a similar average seed number per berry showed very different berry flesh weights: for example, the average berry flesh weight at maturity of cv. Courbu and cv. Assoued Kere, diverged from 1.20 g to 3.36 g respectively with an average of 3.1 seeds per berry weighting 0.2 g. However, the genotypes with berries weighting more than 2 g at *véraison* (4 g or more at maturity) always showed a seed weight per berry upper or equal to 0.1 g (Figure 3B). If the intra-genotypic variation was considered using individual data from ten berries per genotype on the representative sample of 26 genotypes (Table S1), a significant correlation between berry flesh weight variation and seed weight or the seed number variation per berry were observed (Table 2).

The number of carpels was not significantly correlated with the average ovary width and volume or with the average berry flesh weight. Like for seed weight, a threshold effect



**Figure 4.** Distribution of the berry flesh weight (A) at *véraison* and (B) at maturity in the three sub-populations. Table grape varieties originating from Eastern Europe (TE) are in gray, wine grape varieties originating from the Western Europe (WW) are in dark-gray and wine grape varieties originating from Eastern Europe (WE) are in black.



**Figure 5.** Clustering using the Ward method of cultivated *V. vinifera* genotypes based on berry flesh weight, berry length and berry width at *véraison*. K is the number of groups. The number of genotypes (n) per Group for K= 4 is indicated for each of Groups A, B, C, D (from the smallest to the largest berries).

was observed, with the genotypes having large berries (more than 2 g at *véraison*) always harboring large ovaries (upper or equal to 0.1 cm wide; Figure 3C).

Finally, the cell number and volume were measured in ovaries and in berries at *véraison*. Indeed, the berries have their final number of cells at *véraison* (Ojeda et al., 1999) and the final berry size was shown to be mainly determined before this stage in the present study. These traits were variable between the genotypes (Table 1; Figure 2D and E), respectively with a factor of 3.7, 2.3, 1.9 and 4.6. However, there was not significant correlation between berry flesh weight variation and these ovarian and berry traits (data not shown).

### **Clustering of the genotypes to reveal mechanism of berry size variation in grapevine**

The distribution of berry flesh weight in each of the three sub-populations was not superimposed at *véraison* or at maturity (Figure 4). The mean berry flesh weight was significantly different between the three sub-populations (t test between all pairs of sub-populations P-value < 0.05; data not shown). Different mechanisms involved in berry size variation that would have been selected independently in each of these sub-populations could explain these observations. In order to test this hypothesis, the correlations between the different parameters were calculated within the three sub-populations. No correlation was found between any of the different sub-traits (ovary volume, carpel number, ovary cell number and volume, berry flesh cell number and volume, and seed number and weight) and berry flesh weight variation in any of the three sub-populations (data not shown). These results would indicate that the selection for berry size in grapevine is independent of the evolutionary forces that shaped the cultivated *V. vinifera* germplasm into three sub-populations.

The Ward's method where genotypes were clustered using several phenotypic data (berry flesh weight, berry length and berry width at *véraison* together) was used in order to group the genotypes according to their berry weight and volume and without any *a priori* on their origin or genetic structure. The dendrogram obtained after clustering is shown in Figure 5. The distribution in four groups (A, B, C, D) was adopted as it allowed the best balance in term of numbers of individuals in each class (Figure 5), and a significantly different berry flesh weight at *véraison* between them (Table 3). There was however still an asymmetric distribution of the genotypes between groups, Groups A and B containing the majority of the genotypes (respectively 32% and 50%) whereas Groups C and D represented together 18% of all the genotypes (Table 4). Group A that was made of the genotypes with the smallest berries

**Table 3.** Average values and standard deviation of the main parameters measured on ovaries and berries at *véraison* stage from cultivated sample and a wild genotype.

	Group A		Group B		Group C		Group D		Wild grape ( <i>V. vinifera</i> subsp. <i>sylvestris</i> )
	Average <sup>§</sup>	Standard deviation	Average <sup>§</sup>	Standard deviation	Average <sup>§</sup>	Standard deviation	Average <sup>§</sup>	Standard deviation	Average
<b>OVARIES</b>									
Carpel number	2.0 a	0.1	2.1 a	0.1	2.2 a	0.2	2.2 a	0.3	2.0
% of genotypes showing berries with three carpels	15	—	28	—	67	—	63	—	0
Length (cm)	0.12 a	0.01	0.13 b	0.02	0.14 b	0.02	0.14 ab	0.03	0.1
Width (cm)	0.10 a	0.02	0.10 ab	0.02	0.12 bc	0.02	0.13 c	0.01	0.08
Volume (cm <sup>3</sup> )	0.0006 a	0.0002	0.0008 ab	0.0004	0.0010 b	0.0004	0.0013 b	0.0003	0.0003
Shape estimator	1.3 a	0.2	1.3 a	0.1	1.3 a	0.2	1.2 a	0.1	1.3
Outer mesocarp cell density (cell.µm <sup>-2</sup> )*	0.0098 a	0.0028	0.0099 a	0.0020	0.0100 a	0.0031	0.0088 a	0.0020	0.0051
Inner mesocarp cell density (cell.µm <sup>-2</sup> )*	0.0092 a	0.0014	0.0088 a	0.0014	0.0115 a	0.0008	0.0096 a	0.0013	0.0067
Total mesocarp cell density (cell.µm <sup>-2</sup> )*	0.0090 a	0.0020	0.0087 a	0.0018	0.0109 a	0.0019	0.0092 a	0.0010	0.0059
Indicator of cell number*	1202 a	255	1742 b	447	1799 ab	1031	1886 b	432	970
<b>BERRIES</b>									
Flesh weight (g)	<b>0.72 a</b>	0.13	<b>1.15 b</b>	0.19	<b>1.97 c</b>	0.39	<b>3.27 d</b>	0.30	0.29
Length (cm)	1.2 a	0.1	1.4 b	0.1	1.8 c	0.3	2.2 d	0.2	0.8
Width (cm)	1.2 a	0.1	1.3 b	0.1	1.5 c	0.2	1.7 d	0.2	0.8
Volume (cm <sup>3</sup> )	0.8 a	0.2	1.3 b	0.2	2.4 c	0.6	3.4 d	0.6	0.2
Shape estimator	1.0 a	0.1	1.1 b	0.1	1.2 c	0.2	1.3 d	0.2	1.0
% of genotypes with non-spherical berries	17	—	29	—	45	—	80	—	-
Seed number	1.9 a	0.6	1.9 a	0.7	1.9 a	0.8	1.9 a	0.6	1.0
Seed weight (g)	0.10 a	0.03	0.10 ab	0.04	0.12 ab	0.05	0.14 b	0.04	0.06
Outer mesocarp cell density (cell.µm <sup>-2</sup> )*	0.00081 a	0.00021	0.00057 a	0.00011	0.00044 b	0.00001	0.00035 b	0.00013	0.00081
Inner mesocarp cell density (cell.µm <sup>-2</sup> )*	0.00005 a	0.00002	0.00004 a	0.00001	0.00002 b	0.00001	0.00004 b	0.00001	0.00008
Total mesocarp cell density (cell.µm <sup>-2</sup> )*	0.00043 a	0.00010	0.00030 a	0.00006	0.00023 b	0.00001	0.00020 b	0.00007	0.00044
Indicator of cell number*	1870 a	282	2170 ab	460	1961 ab	226	2521 b	304	1729
Indicator of cell expansion rate <sup>§</sup>	0.052a	0.026	0.037a	0.012	0.022b	0.004	0.019b	0.009	0.075
Indicator of cell division rate <sup>ε</sup>	1.7a	0.7	0.3a	0.3	1.4a	0.9	1.4a	0.6	1.8
Number of genotypes	<b>75 (8*)</b>		<b>118 (11*)</b>		<b>32 (2*)</b>		<b>10 (4*)</b>		<b>1(1*)</b>

\* Number of genotypes from the corresponding Group used for microscopic analysis

§ For each parameter, the averages that are followed by the same letter do not differ significantly (p-value <0.05 after Bonferroni correction)

§ Ratio between the total mesocarp cell density in berries and in ovaries

ε Ratio between the indicators of cell number in berries and in ovaries.

contained mainly genotypes from the sub-populations WW (64%) and WE (28%) used for wine industry whereas Group D, containing grapes with the largest berries, only contained genotypes from the sub-population TE (100%) used as Table grapes. Groups B and C contained a mix of the three sub-populations, with a higher proportion of TE individuals in Group C than in Group B. The sample distribution reflects the representation of European wine varieties in the French National Grapevine Germplasm Collection, which has a majority of genotypes producing small to medium berries (Table 4; Cordonnier, 1976; Le Cunff et al., 2008). The comparison of the distribution of the genotypes in the three sub-populations with their distribution in the four clusters based on berry size parameters showed that these two groupings are not independent (chi-squared contingency table test P-value < 0.05; data not shown). This unbalanced distribution of the genotypes according to their berry size in the French National Grapevine Germplasm Collection was found again in the sample of 26 genotypes used for ovarian and berry cell phenotyping, these 26 genotypes having been chosen at the beginning of the study, in absence of in depth phenotypic analysis, to be representative of the three genotypic sub-populations (Table 3).

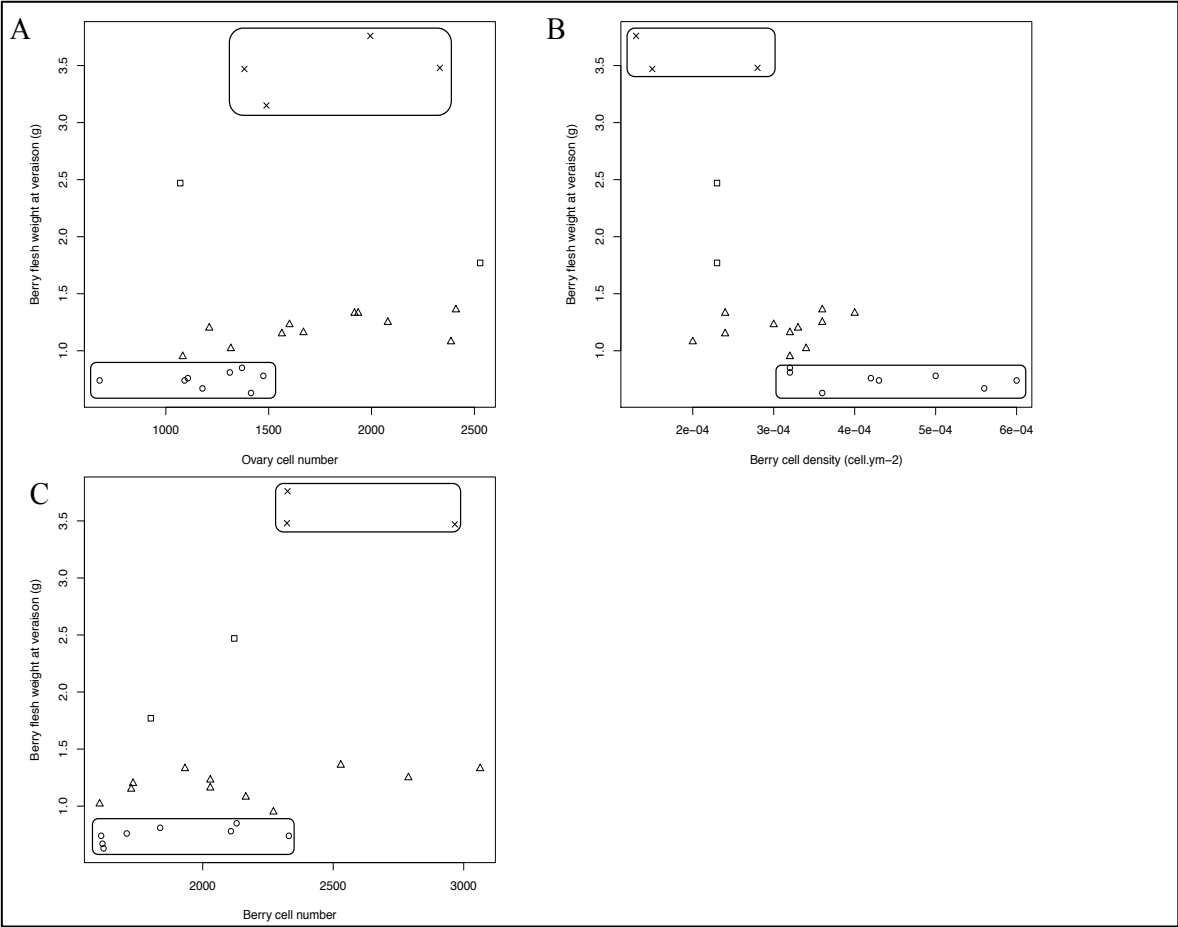
### **Are there different mechanisms underlying berry size variation in grapevine?**

All the parameters measured on berries and ovaries were analyzed within and between genotypes from Groups A, B, C and D (Table 3). Significant differences between some to all Groups were found for all the parameters except ovary cell volume, ovary shape estimator (t test P-value < 0.05; data not shown), carpel number and seed number (generalized linear model P-value < 0.05; data not shown). With regards to the ovary, its volume differed significantly between Groups, with berries in average twice larger in Group D than in Group A (Table 3; Figure 2B). This difference in ovary volume was mainly due to a difference in ovary width for Group D whereas for Group C, both ovary width and length participated (Table 3). Regarding the cell-related parameters measured on ovaries, the cell volume estimated through the mesocarp cell density was not significantly different among the four groups whereas the cell number was significantly different between Group A and Groups B and D (Table 3; Figure 6A; Figure 2D). Although the carpel number did not seem to be directly involved in the ovary volume or berry volume variation (Table 3), the number of genotypes which possessed ovaries with three carpels was higher in Groups with genotypes producing the largest berries than in the others (67% and 63% respectively for Groups C and D; Table 3).

**Table 4.** Correspondence between the three sub-populations based on the global analysis of the genetic diversity of cultivated grapevine (WW, WE and TE) and groups of genotypes based on phenotypic data of berry size (A, B, C and D, from the smallest to the largest berries).

Groups	Number of genotypes	% of total genotypes	% of Vassal collection*	Repartition of genotypes in the 3 sub-populations					
				WW		WE		TE	
				number	%	number	%	number	%
A	75	32	25	48	57	21	26	6	9
B	118	50	48	36	42	50	63	32	46
C	32	14	21	1	1	9	11	22	31
D	10	4	6	0	0	0	0	10	14
Total	235	100	100	85	100	80	100	70	100
* French National Grapevine Germplasm Collection, Domaine de Vassal, Montpellier, France.									

**Figure 6.** Correlation between berry flesh weight at *véraison* and (A) ovary cell number, (B) berry cell density and (C) berry cell number, on a sample of 26 genotypes from the four Groups. (circle: Group A, triangle: Group B, plus: Group C and cross: Group D). The berry cell volume is approached by the berry cell density (see Materials and Methods).



As expected, the seed number did not vary among Groups whereas the seed weight was significantly different between Groups A and D (Table 3). Groups C and D had significantly larger berry flesh cells than Groups A and B (Figure 6B; Figure 2E) in both outer and inner mesocarps and this differentiation clearly started after anthesis as shown both by the significant differences between the two development stages in cell density and in their expansion rate (Table 3). In addition, the number of berry flesh cells was still significantly smaller in Groups A than in Group D (Table 3; Figure 6C). However, in this case the cell division rate in mesocarp between the anthesis and the *véraison* stages is not significantly different between the four groups (Table 3). The ovary shape estimator did not differ significantly between groups (Table 3) and was not correlated to the final berry shape. Opposite, the berry shape estimator at *véraison* differed significantly between the four groups and was highly correlated to the berry shape at maturity (Figure 2F). Interestingly, the genotypes producing non-spherical berries were more frequent in Groups with large berries (45% and 80% of the genotypes for Group C and D respectively; Table 3). In addition, all groups, except Group A, comprised several genotypes producing berries with shapes different from round or ellipsoid (Figure 2F). For four of the parameters with a normal distribution, for which data were collected for three to ten individual berries in 26 individuals (berry weight, seed weight, berry cell density, ovary cell density), an anova analysis, taking into account the intra-genotypic variation (Table S2) confirmed the results presented above (Table S3).

To summarize, the berry size variation originated from variation in both ovary and berry traits (Table 3). In Group A, which contained the genotypes producing the smallest berries, berry size could be explained both by few ovarian cells and probably a cell growth limited after anthesis, resulting in berries at *véraison* with fewer and smaller cells. Group B was characterized by a higher number of cells in the ovaries, which was more or less maintained during berry development, depending on the genotypes. However, the size of the berry cells was not significantly different with those of Group A at *véraison*. The number of cells at anthesis seemed therefore to be the factors explaining the berry size variation between Group A and B. The cell volume in berries at *véraison* was clearly the factor that differentiated Groups A and B from C and D. In Group C, the number of cells in both ovaries and berries remained variable whereas it was significantly higher in Group D than in Group A. Finally, Group D, which contained the genotypes with the biggest berries, had the largest ovaries at anthesis, containing lots of large cells, resulting into very large berries with lots of large cells.

**Preliminary data on the differences between the wild and cultivated *V. vinifera* berries**

Finally, in order to approach the difference between wild and cultivated *V. vinifera* berries, the same ovarian and berry traits were measured in a female wild genotype (Table S1). The data obtained are reported in Table 3. All the ovarian and berry sub-traits, except the number of carpel, were in average lower in the wild genotype than in the cultivated genotypes but still in the range of variation of genotypes from Group A (Table 1; Table 3). The carpel number in the ovary was invariable and equal to two. These preliminary results would suggest that wild ovaries and berries do not differ much from some cultivated genotypes present in Group A. However, at *véraison*, the berry flesh weight and volumes that were measured from the wild accession were clearly out of range of the cultivated accessions (Table 3). This result would be explained by the fact that in Group A the genotypes showing the smallest number of berry cells were not the ones with the smallest berry cell size (data not shown).

**Discussion**

The correlation between berry weight at *véraison* and at maturity in a large collection of diverse grapevine genotypes observed in this study indicated that the main determinants of the genetic variation for berry size occurs before *véraison*, that is to say during a stage of intense cell divisions associated with some cell extension in the grapevine fruit (Ojeda et al., 1999). After *véraison*, fruit growth is only related to cell extension in grapevine and therefore the contribution of this late mechanism in berry size genetic variation seems to be rather low.

Indicators of the cell number and size in ovaries and berries were produced from histological observations in 26 genotypes. Although the sampling was carefully made to be representative of the genetic variability of the whole sample, this number of genotypes was a bit low a study of the genetic variation of a trait and the C and D Groups appeared to be under-representated. With this sampling, cell division before anthesis and cell expansion after anthesis were found to be involved in flesh weight variation between varieties in grapevine. Indeed, Coombe (1973) showed in grapevine that 17 doublings of cell number were occurring in ovary before anthesis compared to only two doublings after, which suggested that cell number in mature fruit was mainly determined before anthesis. In contrast, cell expansion increased fruit size by a 300-fold factor and more after anthesis (Coombe, 1976) and was the greatest contribution to the total expansion of the fruit. This is in agreement with the present results showing that cell volume was stable between varieties at anthesis but showed variation at maturity where it was correlated with final fruit size. The increase of berry volume *via* flesh cell expansion was shown influenced by the cell wall plasticity and the accumulation of water



and sugars (Coombe, 1976), which is in turn certainly influenced by the number of vascular bundles. Interestingly, some preliminary results indicate that ovaries from Group D would present more vascular bundles than the ovaries from Group A (data not shown). Ovarian cell number and ovary vascular bundle number were already shown to be involved in sorghum kernel size variation by Yang et al (2009) and Jackson and Coombe (1960) also demonstrated that cell number and volume at anthesis were the most important factors contributing to fruit size variation in apricot. In tomato fruit, the role of cell number into fruit variation was also demonstrated to occur at anthesis or before anthesis, increasing the range of fruit variation in combination with other traits such as berry cell size and carpel number (Bertin et al., 2003; Cheniclet et al., 2005; Baldet et al., 2006; Cong et al., 2008).

Each of these components may in turn be influenced by growth regulators (Ollat et al., 2002; Coombe, 1976). In the present study, a positive correlation was found between the seed number and weight and the berry weight at the intra-genotypic level, which was in agreement with previous results (Coombe, 1960; Ollat, 2002; Dai et al., 2009a). The seed effect on berry growth was mainly related to growth regulator supply (Coombe, 1973). Ojeda et al (1999) suggested that seed growth might have a positive effect on cell mitosis in berry flesh development and Dai et al. (2009a) showed that the seed number affects berry growth especially during the first growth phase during which cell divisions occurs. In tomato, auxins produced by the seeds and/or surrounding fruit tissues are also known to trigger or accelerate fruit growth by cell expansion (Gillaspy et al., 1993; Lemaire-Chamley et al., 2005). The seed content could therefore influence the berry size through hormonal mechanisms, more seeds or larger seeds producing more hormones than fewer of smaller ones. Another hypothesis would be that in large berries, seeds would find a less limited space to grow. However, although the genotypes with the largest berries always had the highest seed weight, the absence of overall correlation between seed number or seed weight and berry flesh weight at the inter-genotypic level in cultivated grapevine suggested that this mechanism does not account much for berry size variation. The observed threshold effect would have to be tested with a larger sample of genotypes carrying large berries.

If the fruit size in cultivated tomato can vary by a factor of 200-fold (Tanksley, 2004; Cheniclet et al., 2005), only a 10-fold variation was observed in the highly diverse set of 254 cultivated grapevine accessions phenotyped in the present work. This difference could be explained in part by the number of carpels that is highly variable in tomato and has been demonstrated to have a strong influence in tomato fruit weight variation (Cong et al., 2008). This factor was proposed to be under selection pressure after the increase of ovarian cell

number during the history of breeding for fruit size in tomato (Tanksley, 2004). In grapevine, this trait was shown in the present study to be much less variable than in tomato as it varied between two and three carpels per ovaries *versus* two to more than 20 in tomato (Cheniclet et al., 2005). However, although this trait does not seem to contribute to the overall berry size variation, varieties from Groups C and D which produced the largest berries contained more genotypes with three carpel than Groups A and B.

The present study also suggested that, like in tomato (Tanksley, 2004), the increase of berry weight in grapevine is correlated with the increase of berry shape diversity. This could be due either to loci having pleiotropic effects on fruit shape and fruit size or to parallel but independent selections for both traits in order to satisfy consumer demands. Indeed, the wild grapevine and tomato fruits are invariably round and small (Tanksley, 2004; This et al., 2006) whereas cultivated fruits showed a large range of size and shape. In favor of the first hypothesis, several genes involved in fruit size variation have a pleiotropic effect on fruit shape in tomato (see for instance *FASCIATED* and *LC* which control the carpel number; Tankley, 2004). Moreover, the timing of divisions in different parts of flesh, but also the direction of divisions (periclinal, anticlinal, random), and the interaction of both with the timing of cell expansion, influence the final fruit shape (Coombe, 1976) and could explain the correlation between fruit size and shape diversities. The fact that the final berry shape was not predictable at the ovary stage may either be due to the accuracy of the method used for its measure or to the fact that this trait is determined after fertilization.

None of the measured traits variation could be superimposed with the evolutionary history that lead to the genetic structure of the grapevine cultivated gene pool in three sub-populations WW, WE and TE. It was finally the clustering of genotypes in four Groups (A, B, C, D) based on berry size traits (berry weight, length and width) that allowed revealing the important mechanisms involved in berry size variation in cultivated grapevine. The transition of berry size between the four Groups of varieties could be explained by a progressive selection on main traits: the berry cell number determined by the cell number in the ovaries, the intensity of cell divisions after anthesis and the berry cell extension after anthesis. A natural mutant of *V. vinifera* cv. Ugni Blanc, which produces fleshless berries (Fernandez et al., 2006b), showed also fewer and smaller flesh cell than those from Ugni Blanc wild-type (Fernandez et al., 2006a), indicating that the mutation in one locus is enough to explain such a transition. The most likely evolutionary scenario that would fit to these observations is that human gradually selected mutations associated with an increase of fruit size, progressively leading to the present day cultivars showing large range of size. Such scenario was already

mentioned for tomato fruit history (MacArthur and Butler, 1938; Banerjee and Kalloo, 1989). The factor explaining the transition between the wild and the cultivated genotypes regarding to berry size is unclear from these preliminary data, as none of the measured parameters seemed clearly out of the range of variation of the genotypes carrying small berries (Table 2). However, it seems to be the combination of both few cells and very little cells (found separately in cultivated genotypes) that is responsible of the very small wild berries. The grapevine domestication would have mainly concern other traits like the change from dioecy to hermaphroditism, and perhaps an increase of bunch size and density (This et al., 2006) but not much berry size which is not a favourable trait for wine cultivars that could explain the results. The phenotyping of a representative set of wild female *V. vinifera* representative of the wild diversity would be necessary to better understand the historical processes of selection for berry size.

Berry size is also influenced by factors such as berry location within the cluster, the number of berries per cluster, the plant source / skin ratio (Ollat, 2002; Dai et al., 2009b). In this preliminary study, the influence of the two first parameters on berry size variation was taken into account by randomly sampling berries within several clusters from several plants.

It will be important in future studies to design specific experiments for a better estimation of the interaction between genetic and environmental factors involved in berry size variation. In this study however, there was a high correlation between the average berry weight at *véraison* of 18 control genotypes measured two successive years. This study pointed the *véraison* stage as the most interesting for the study of the berry size variation, as the berry size at *véraison* and maturity are highly correlated while sampling at *véraison* is much more reliable than at maturity. Indeed, the transition stage between maturity and over-maturity where berry size can start to decrease (McCarthy, 1999) is difficult to detect visually, especially for such a large collection of diverse genotypes.

## Authors' contributions

CH designed the experiments, phenotyped the collection for ovary and berry traits, realized and analyzed histological cuts of ovaries and berries, analyzed the result, realized statistical analysis, draft and corrected the manuscript. M-LM realized the clustering using Ward's method, helped in statistical analysis. SDN, LT and LCL generated the panels of genotypes. DF analyzed histological cuts of ovaries. GC helped for realizing of histological cuts of berries and ovaries. LT, PT and SL helped for the design of the experiment and for phenotyping. A-FAB design the experiment, supervised it, drafted and corrected the manuscript.

## Acknowledgements

We warmly thank Mireille Dessup and all the team of the INRA Domaine de Vassal (Marseillan-Plage, France) for technical assistance, Jacques Escoute, Marc Lartaud, Jean-Luc Verdeil, and Fabienne Montes (Histology and Plant Cell Imaging Platform, CIRAD, Montpellier, France; <http://phiv.cirad.fr/>) for helping in microscopical analysis, Cécile Guichard for made-home script for data analysis and Amidou N'Diaye for helping during phenotyping campaigns.

## References

- Baldet P, Hernould M, Laporte F, Mounet F, Just D, Mouras A, Chevalier C, Rothan C.** 2006. The expression of cell proliferation-related genes in early developing flowers is affected by fruit load reduction in tomato plants. *Journal Experimental of Botany* **57**: 961-970.
- Banerjee MK and Kalloo G.** 1989. The inheritance of earliness and fruit weight in crosses between cultivated tomatoes and two species of *Lycopersicon*. *Plant Breeding* **102**:148-152.
- Bertin N, Borel C, Brunel B, Cheniclet C, Causse M.** 2003. Do genetic makeup and growth manipulation affect tomato fruit size by cell number, or cell size and DNA endoreduplication? *Annals of Botany* **92**: 415-424.
- Boursiquot J-M, Dessup M, Rennes C.** 1995. Distribution des principaux caractères phénologiques, agronomiques et technologiques chez *Vitis vinifera* L. *Vitis* **34**: 31-35.
- Cabezas JA, Cervera MT, Ruiz-Garcia L, Carreno J, Martinez-Zapater JM.** 2006. A genetic analysis of seed and berry weight in grapevine. *Genome* **49**: 1572-1585.
- Cheniclet C, Rong WY, Causse Mathilde, Frangne N, Bolling L, Carde JP, Renaudin JP.** 2005. Cell expansion and endoreduplication show a large genetic variability in pericarp and contribute strongly to tomato fruit growth. *Plant Physiology* **139**: 1984-1994.

- Cong B, Lui J, Tanksley SD.** 2002. Natural alleles at a tomato fruit size quantitative trait locus differ by heterochronic regulatory mutations. *Proceedings of the National Academy of Sciences* **99**: 13606-13611.
- Cong B, Barrero LS, Tanksley SD.** 2008. Regulatory change in YABBY-like transcription factor led to evolution of extreme fruit size during tomato domestication. *Nature Genetics* **40**: 800-804.
- Coombe BG.** 1960. Relationship of growth and development to changes in sugars, auxins, and gibberellins in fruit of seeded and seedless varieties of *Vitis vinifera*. 1960. *Plant Physiology* **35**: 241-250.
- Coombe BG.** 1973. The regulation of set and development of the grape berry. *Acta Horticulturae* **34**: 261-271.
- Coombe BG.** 1976. The development of fleshy fruits. *Annual review of plant physiology* **27**: 507-528.
- Cordonnier E.** 1976: Qualité de la vendange et méthodologie de la sélection viticole. *Progrès Agricole et Viticole* **93**: 760-762.
- Dai ZW, Genard M, Li SH, Vivin P.** 2009a. Analyzing the functional association among seed traits, berry growth and chemical composition in Cabernet-Sauvignon berry (*Vitis vinifera* L.) using a mathematical growth function. *Journal International des Sciences de la Vigne et du Vin* **43**: 35-44.
- Dai ZW, Vivin P, Robert T, Milin S, Li SH, Genard M.** 2009b. Model-based analysis of sugar accumulation in response to source-sink ratio and water supply in grape (*Vitis vinifera*) berries. *Functional Plant Biology* **36**: 527-540.
- Doligez A, Bouquet A, Danglot Y, Lahogue F, Riaz S, Meredith CP, Edwards KL, This P.** 2002. Genetic mapping of grapevine (*Vitis vinifera* L.) applied to the detection of QTL for seedlessness and berry weight. *Theoretical and applied genetics* **105**: 780-795.
- Fanizza G, Lamaj F, Costantini L, Chaabane R, Grando MS.** 2005. QTL analysis for fruit yield components in table grapes (*Vitis vinifera*). *Theoretical and applied genetics* **111**: 658-664.
- Fernandez L, Pradal M, Lopez G, Berud F, Romieu C, Torregrosa L.** 2006a. Berry size variability in *Vitis vinifera* L. *Vitis* **45**: 53-55.
- Fernandez L, Romieu C, Moing A, Bouquet A, Maucourt M, Thomas MR, Torregrosa L.** 2006b. The Grapevine fleshless berry Mutation. A Unique Genotype to Investigate Differences between Fleshy and Non Fleshy Fruit. *Plant Physiologie* **140**: 537-547.
- Frary A, Nesbitt TC, Grandillo S, Knaap E, Cong B, Liu J, Meller J, Elber R, Alpert KB, Tanksley SD.** 2000. *FW2.2*: a quantitative trait-locus key to the evolution of tomato fruit size. *Science* **289**: 85-88.
- Gillaspy G, Ben-David H, Gruissem W.** 1993. Fruits: a developmental perspective. *Plant Cell* **5**: 1439-1451.

**Jackson DI and Coombe BG.** 1966. The growth of apricot fruit. I Morphological changes during development and the effects of various tree factors. *Australian Journal of Agricultural Research* 17: 465-477.

**Houel C, Bounon R, Chaïb J, Guichard C, Peros J-P, N'Diaye A, Dereeper A, Canaguier A, Lacombe T, Bacilieri R, Le Paslier M-C, Vernerey M-S, Coriton O, Brunel D, This P, Torregrosa L, Adam-Blondon A-F.** 2010. High level of homozygosity and signatures of selection in the region of the fleshless berry locus in grapevine. *BMC Plant Biology* 10: 284.

**Laucou V, Lacombe T, Dechesne F, Siret R, Bruno J-P, Dessup M, Dessup T, Ortigosa P, Parra P, Roux C, Santoni S, Vares D, Peros J-P, Boursiquot J-M, This P.** 2011. High throughput analysis of grape genetic diversity as a tool for germplasm collection management. *Theoretical and applied genetics* 122: 1233-45.

**Le Cunff L, Fournier-Level A, Laucou V, Vezzulli S, Lacombe T, Adam-Blondon A-F, Boursiquot J-M, This P.** 2008. Construction of nested core collections to optimize the exploitation of natural diversity in *Vitis vinifera* L. *subsp sativa*. *BMC Plant Biology* 8: 31.

**Lemaire-Chamley M, Petit J, Garcia V, Just D, Baldet P, Germain V, Fagard M, Mouassite M, Cheniclet C, Rothan C.** 2005. Changes in transcriptional profiles are associated with early fruit tissue specialization in tomato. *Plant Physiologie* 139: 750-769.

**Lorenz, D.H, Eichhorn K.W, Blei-holder H, Klose R, Meier H, Weber E.** 1994. Phänologische Entwicklungsstadien der Weinrebe (*Vitis vinifera* L. *ssp. vinifera*). *Wein-Wissenschaft* 49: 66-70.

**MacArthur JW and Butler L.** 1938. Size inheritance and geometric growth processes in the tomato fruit. *Genetics* 23: 253-268.

**Mccarthy MG.** 1999. Weight loss from ripening berries of Shiraz grapevines (*Vitis vinifera* L. cv Shiraz). *Australian Journal of Agricultural Research* 5: 10-16.

**Mejia N, Soto B, Guerrero M, Casanueva X, Miccona M, Houel C, Le Cunff L, Boursiquot JM, Hinrichsen P, Adam-Blondon AF.** 2011. Molecular, genetic and transcriptional evidence for a role of VvAGL11 in stenospermocarpic seedlessness in grapevine. *BMC plant biology* 11: 57.

**Mejia N, Gebauer M, Munoz L, Hewstone N, Munoz C, Hinrichsen P.** 2007. Identification of QTL for seedlessness, berry size, and ripening date in a seedless x seedless table grape progeny. *American Journal of Enology and Viticulture* 58: 4.

**Ojeda H, Deloire A, Carbonneau A, Ageorges A, Romieux C.** 1999. Berry development of grapevines: relations between the growth of berries and their DNA content indicate cell multiplication and enlargement. *Vitis* 38: 145-150.

**Ollat N, Diakou-Verdin P, Carde J-P, Barrieu F, Gaudillère J-P, Moing A.** 2002. Grape berry development: a review. *Journal international des sciences de la vigne et du vin* 36: 109-131.

**Pritchard JK, Stephens M, Donnelly P.** 2000. Inference of population structure using multilocus genotype data. *Genetics* **155**: 945-959.

**Purugganan MD and Fuller DQ.** 2009. The nature of selection during plant domestication. *Nature* **457**: 843-848.

**Tanksley SD.** 2004. The Genetic, Developmental, and Molecular Bases of Fruit Size and Shape Variation in Tomato. *Plant Cell* **16**: 181-189.

**This P, Lacombe T, Thomas MR.** 2006. Historical origins and genetic diversity of wine grapes. *Trends Genetics* **22**: 511-519.

**Yang Z, Oosterom EJ, Jordan DR, Hammer GL.** 2009. Pre-anthesis ovary development determines genotypic differences in potential kernel weight in sorghum. *Journal Experimental of Botany* **60**: 1399-1408.





### 3.4. Conclusion du chapitre

Le phénotypage d'une collection de plus de 300 génotypes sur des caractères de l'ovaire et de la baie reste une opération lourde et saisonnière, le moindre aléa peut générer de nombreuses données manquantes. Le phénotypage des baies à véraison que je propose à l'issue de ma thèse comporte le double avantage d'être plus précis en terme d'échantillonnage (stade très facilement identifiable), mais également en terme de limitation de la possibilité d'apparition d'aléas affectant les baies, qui augmentent au cours du temps. Bien que la croissance des baies de l'anthèse à la véraison soit plus sensible aux paramètres extérieurs que durant la maturation, avec l'occurrence d'altérations de la croissance de la baie irréversibles (Ollat, 2002), la corrélation entre la taille des baies à véraison et à maturité reste cependant inchangée au sein d'une même année de phénotypage. Il faut noter cependant que l'environnement semble avoir eu très peu d'influence sur le poids de la baie pendant mes deux années de récoltes (résultat sur 18 génotypes contrôles phénotypés sur les deux années). Le phénotypage des 303 génotypes cultivés sur plusieurs années et à des sites différents permettrait d'estimer la composante environnementale imputée à chaque caractère, et l'identification de gènes responsables de la variation de la taille de la baie dans tous les environnements, ou au contraire, permettrait d'identifier des gènes sous forte influence environnementale qui pourraient être utilisés en sélection.

En revanche, nous n'avons pas pu proposer une technique haut-débit pour le phénotypage du nombre et de la taille des cellules du mésocarpe. Dans cette étude, j'ai approché le nombre de cellules de la baie par le nombre de cellules dans une coupe transversale, il serait maintenant important de réaliser des coupes longitudinales pour affiner les comptages et vérifier si le passage des baies de la forme ronde à ovale est dû à un allongement asymétrique des cellules du mésocarpe ou à une augmentation asymétrique du nombre de cellules.

Cette étude a permis de mettre en évidence la complexité des mécanismes à l'origine de la variation de la taille de la baie chez la vigne cultivée, faisant intervenir le nombre de cellules déjà déterminé dans l'ovaire avant l'anthèse et la capacité des cellules du mésocarpe à grossir après l'anthèse. Ces mécanismes sont similaires à ceux expliquant la variation de la taille du fruit chez la tomate. En effet, l'augmentation du volume des cellules du péricarpe après l'anthèse (Cheniclet et al, 2005; Bertin et al, 2003), ainsi que le nombre de cellules dans l'ovaire à l'anthèse (Frary et al, 2000) expliquent en partie la différence de poids final du fruit



**Figure 3.3:** Photo d'une vigne ravagée par les lapins. Les flèches blanches indiquent les branches dépourvues d'inflorescences et de feuilles.

chez la tomate, le nombre de carpelles expliquant également une part importante de la variation. Bien qu'aucune corrélation significative n'ait été détectée entre le nombre de carpelles dans l'ovaire et la taille finale de la baie (poids/volume) chez la vigne, il y a plus de génotypes produisant de grosses baies avec trois carpelles que de génotypes produisant de petites baies. Le rôle du nombre de carpelles dans la variation de la taille de la baie reste donc à éclaircir dans une collection plus importante de variétés à grosses baies.

Enfin, dans cette étude, un seul génotype sauvage a pu être phénotypé pour les caractères de l'ovaire et de la baie, les autres génotypes sauvages présents sur le site du Domaine de Vassal étant soit des mâles, soit trop jeunes pour produire des baies, ou bien encore avaient eu toutes leurs inflorescences dévorées par des lapins (Figure 3.3). Il serait important dans le futur de mieux caractériser la variabilité dans le compartiment sauvage pour en préciser les différences avec le compartiment cultivé et en particulier avec les variétés de cuve.

Dans l'objectif de contribuer à l'identification de gènes impliqués dans la variation de la taille de la baie chez la vigne cultivée, des associations ont été recherchées entre des polymorphismes de séquence de gènes candidats pour la variation de la taille de la baie et la variation de l'ensemble des caractères mesurés à l'exception du nombre et de la taille des cellules qui avaient été estimés sur un trop petit échantillon de génotypes (26). Ce travail est présenté dans le chapitre suivant.



#### **4. Recherche de gènes impliqués dans la variation de la taille de la baie chez *V. vinifera* ssp *sativa*: étude de génétique d'association**



*Photographie d'une grappe de raisins au stade petit pois.*



## 4.1. Introduction

Dans le chapitre précédent, nous avons montré dans une collection de 254 génotypes que la variation de la taille de la baie chez la vigne cultivée était définie précocement par le nombre de cellules de l'ovaire et par la capacité de ces cellules à grossir après l'anthèse. Cependant d'autres paramètres mesurés comme le nombre de carpelles par ovaire, le nombre et la taille des pépins de la baie bien que variables d'un génotype à l'autre ne semblent pas associés à la variation taille de la baie entre les génotypes.

On ne dispose pour l'instant que de quelques études de détection de QTL pour le poids de la baie dans des croisements avec des variétés apyrènes (Doligez et al, 2002; Fanizza et al, 2005; Cabezas et al, 2006; Mejia et al, 2007; Costantini et al, 2008). Ces QTL sont majoritairement différents entre les études et montrent une instabilité en fonction des années de phénotypage. Ces études montrent la difficulté à identifier les déterminants génétiques de la variation de la taille de la baie chez la vigne, en raison de la lourdeur des dispositifs expérimentaux qui seraient nécessaires au contrôle des effets environnementaux, mais également au choix des descendance qui ségrégaient en majorité aussi pour le caractère apyrène, en forte interaction avec le caractère taille de la baie. A l'heure actuelle, aucun gène capable d'expliquer ne serait-ce que partiellement la variabilité observée au niveau du poids des baies n'a encore été identifié.

L'objectif de cette partie de ma thèse a été de rechercher des gènes impliqués dans la variation de la taille de la baie par une autre approche, la génétique d'association, dans le large échantillon de génotypes cultivés préalablement phénotypés, puis de renforcer l'implication de ces gènes dans la variation de la taille de la baie par l'étude de leurs profils transcriptomiques au cours du développement de la fleur et de la baie dans différents génotypes contrastés pour la taille de la baie.

## 4.2. Analyse génétique de la variation de la taille des baies

Cette partie est présentée sous la forme d'un article en préparation en vue d'une publication ultérieure.





## Screening candidate genes for berry traits using association genetics in grapevine (*Vitis vinifera* L.)

Cléa Houel<sup>1</sup>, Aurélie Canaguier<sup>1</sup>, Laurent Torregrosa<sup>2</sup>, Loic Le Cunff<sup>2,3</sup>, Richard Berthomé<sup>1</sup>, Aurélie Bérard<sup>4</sup>, Aurélie Chauveau<sup>4</sup>, Marie-Christine Lepaslier<sup>4</sup>, Roberto Bacilieri<sup>2</sup> and Anne-Françoise Adam-Blondon<sup>1</sup>

<sup>1</sup> Unité Mixte de Recherche de Génomique Végétale, INRA UEVE ERL CNRS, 2 rue Gaston Crémieux, 91057 Evry, France

<sup>2</sup>TGU AGAP – Equipe DAVEM, INRA-Montpellier SupAgro, 2 place Pierre Viala, 34060 Montpellier, France

<sup>3</sup> Unité Mixte Technologiques Géno-Vigne, INRA IFV SupAgro, Domaine de l'Espiguette, 30240 Le Grau du Roi, France

<sup>4</sup> Unité INRA Etude du Polymorphisme des Génomes Végétaux, 2 rue Gaston Crémieux, 91057 Evry, France

### Abstract

**Background:** Fruit size is an important trait in fruit crops. Its variation seems to be determined by common but also specific developmental mechanisms in different fleshy fruits. If the genetic and molecular control of fruit size variation has largely progressed in tomato, no gene involved in the regulation of fruit size variation in grapevine has been identified so far. To screen for genes potentially involved in berry size variation, we developed an association study based on a collection of a 240 genotypes of *Vitis vinifera* cultivars. We analysed the links between ovarian and berry traits with SNP profiles of a range of candidates genes selected within QTL for berry weight or from functional annotations.

**Results:** 83 SNP in twenty-six candidate genes were tested and six associations involving five SNP in three genes, *VvAGL11*, *VvJMJ* and *VvHMGR*, were identified. *VvAGL11* was found only associated with the seed number per berry. This gene encode for a putative Agamous like 11 protein well known for its role in the control of ovule and seed identity in Arabidopsis and Petunia and co-localized with both a major QTL for berry size and seedlessness. In grapevine, the expression pattern of this gene was found consistent with its putative function. Two SNP from *VvHMGR1*, a putative 3-hydroxy-3-methyl-glutaryl-CoA reductase (HMGR) were also found associated with berry length. The HMGR is known to be involved in the synthesis of isoprenoid, a precursor of plant growth regulators like cytokinine, abscisic acid and gibberellins. The expression pattern of this gene during flower and early berry development was different between varieties contrasted for berry shape. Finally, the same SNP in the *VvJMJ* gene was found to be involved both with berry weight and berry length variation,

suggesting a putative pleiotropic effect of this gene on fruit shape and fruit size in the grapevine. *VvJMJ* encodes a putative transcription factor jumonji, which was found to be involved in transcriptional repression and/or chromatine regulation. Study of this expression pattern did not support other information about this function. The real role of these genes on each trait remains to be elucidated.

**Conclusions:** We demonstrated the feasibility to association genetics strategy in grapevine using a selection of candidate genes from QTL as previously study but also candidate genes from functional annotation. We bring most interesting candidates for marker-assisted selection and for further functional functional characterisations, which are time and labor consuming in perennials.

## Background

The genetic and molecular bases of fleshy fruit size variation have been studied in depth in tomato during the last two decades (Tanskley, 2004; Paran and Van der Knaap E, 2007). These studies have shown that three traits, the cell number at anthesis, the cell volume after anthesis and the carpel number could explain the final fruit weight variation within the cultivated compartment (Frary et al, 2000; Cheniclet et al, 2005; Cong et al, 2008). In tomato, several genes involved in a variation of fruit weight and shape have been identified so far (Tankley, 2004; Paran and Van der Knaap, 2007). The first one was *Fw2.2*, a negative regulator of cell division with activity confined largely to the cell division phase of fruit development (Frary et al, 2000; Cong et al, 2002). Natural genetic variation at this locus was shown to change the size of fruit by up to 30% (Frary et al, 2000). Two other loci, *Fasciated* and *Locule-number*, were identified later on as causing changes in fruit size *via* changes in the number of carpels in the flower (Cong et al, 2008; Ranc, 2010) and could explain up to 50% of the variation in fruit size. *Fasciated* encodes for a YABBY-like transcription factor that controls carpel number during flower and/or fruit development and *Locule-number* is so far represented by two Single Nucleotide polymorphisms (SNP) located upstream to the transcription factor *Wuschel* but the functional consequence of these polymorphisms remains unclear. Two major loci that modulate fruit shape but having a minimal effect on fruit size were also identified *Ovate* and *Sun* (Liu et al, 2002; Xiao et al, 2008). The mutation in *Ovate*, encoding for a putative nuclear protein expressed during flower development, was associated with a change from round to elongated or pear-shaped fruit (Liu et al, 2002). *Sun* encodes a

member of the IQ67 domain-containing family and an increase in the level of transcription of this gene in flowers and during early fruit development is associated with an elongated fruit shape. All these genes were found to be expressed during flower and/or early fruit development, showing that the final fruit size was determined early in tomato.

In grapevine, like in tomato, there is a large variation of fruit weight in cultivated compartment. Indeed, some wine grape varieties produce berries weighting less than 1 g while table grape varieties produce berries weighing as much as 10 g (Boursiquot et al, 1995; Houel et al, submitted). Like in tomato, the final berry size is determined early during cell division before anthesis and cell expansion after anthesis, these two traits were found to be the major determinants of flesh weight variation between varieties in grapevine (Houel et al, submitted). Conversely, the carpel number per ovary although variable between genotypes did not strongly correlated with the final berry size. In addition, the increase of berry weight in grapevine was also shown to be correlated with the increase of berry shape diversity (estimated through berry length and width; Houel et al, submitted). This could be due either to loci having pleiotropic effects on fruit shape and fruit size or to parallel but independent selections of different genes involved in both traits. Finally, seed number and weight appeared also to influence berry weight, but at intra-genotypic level (Coombe, 1960; Ollat, 2002; Houel et al, submitted) certainly in relation to growth regulator supply (Coombe, 1973).

The genetic analysis of grapevine berry size variation was proved to be however more difficult in grapevine than in tomato, due to several limiting factors such as the long biological cycle of the plant and to the large field area usually required for plant growth. To date, only few Quantitative trait loci (QTL) were published for berry weight (Doligez et al, 2002; Fanizza et al, 2005; Cabezas et al, 2006; Mejia et al, 2007; Costantini et al, 2008; Mejia et al, 2011). However these detections were nearly always performed using populations segregating for seedlessness and there is a strong negative correlation between the two traits (Doligez et al, 2002; Cabazas et al, 2006; Mejia et al, 2007; Constantini et al, 2008; Mejia et al, 2011). These studies identified 12 QTL for berry weight of which only one was common to five of the six studies (all but Fanizza et al, 2005) and that co-localized with the major QTL of seedlessness. No gene involved in berry size variation in grapevine was identified so far. However, recent genetic and transcriptomic analysis pointed the *VvAGL11* gene that co-localise with the major QTL involved both in seedlessness and berry size as a serious functional candidate (Mejia et al, 2011). This gene encode for a putative Agamous like 11 protein well known for its role in the control of ovule and seed identity in Arabidopsis and

Petunia. This work and two others (Fournier-Level et al 2009; Emanuelli et al, 2010), show the interest to combine comparative functional genomics and QTL based and association genetics based genetic approaches to progress in deciphering the molecular control of trait variation. The first study demonstrated that five polymorphisms in *VvMybA* genes accounted for 84% of berry anthocyanin content in a natural population of 141 individuals (Fournier-Level et al 2009). The second study revealed three SNP in *VvDXS* to be significantly associated with muscat-flavored varieties in a population of 148 grape varieties and the unique mutations identifying the muscat-like aromatic mutants (Emanuelli et al, 2010). Finally, a natural mutant showing a heavily reduced development of the flesh is also under progress but the underlying gene (*Fleshless berry* or *Flb*) has not yet been identified (Fernandez et al, 2006a and 2006b; Chaïb et al, 2010; Houel et al 2010).

The aim of this work was to identify genes involved in berry size variation in cultivated *Vitis vinifera* L. (*V. vinifera*) using a association genetics based approach for the identification of genes involved in berry size variation in cultivated *V. vinifera*. It aimed at correlating ovarian and berry traits with SNP from candidate genes within QTL for berry weight, from knowledge in model species such as tomato and Arabidopsis, but also from transcriptomic analysis between a *Flb* mutant and its wild type (Fernandez et al, 2007).

## Material and Methods

### Plant material and phenotyping

All the cultivated *V. vinifera* genotypes used in the present study are listed in Table S1 and are maintained in the French National Grapevine Germplasm Collection (Domain de Vassal, Montpellier, France; <http://www1.montpellier.inra.fr/vassal/>), with the exception of the inbred line INRA Colmar line PN40024 (PN40024), which is maintained at INRA Colmar, France. Twenty-six genotypes corresponding to a sample maximizing the genetic diversity of the cultivated *V. vinifera* compartment (Le Cunff et al, 2008; Houel et al, 2010) and the PN40024 as a control (Jaillon et al 2007) were used for SNP discovery in candidate genes for berry size variation. 223 other cultivated *V. vinifera* were chosen to be representative of the three genetic pools identify using Structure Software (Pritchard et al, 2000) with 20 SSR markers (Laucou et al, 2011; Bacilieri et al, unpublished results). They corresponded to 80 wine cultivars originating from Eastern Europe (WE), 81 wine cultivars originating from the Western Europe (WW) and 62 table cultivars originating from Eastern Europe (TE) and they

represented 80% of the SSR genetic diversity of the French National Grapevine Germplasm Collection. Eight non-related genotypes were also added in order to improve the sampling the phenotypic diversity for berry size. Finally, this sample of 231 genotypes represented well the berry size diversity of cultivated *V. Vinifera* (Houel et al, submitted) and was used for genotyping experiments with SNP from candidate genes for berry size together with nine out of the 26 genotypes used for SNP discovery which produced enough flowers and berries to be phenotyped. The *V. vinifera* cv. Ugni Blanc, its clone which present a fleshless berry phenotype (Fernandez et al, 2006a), and two out of 231 genotypes contrasted for berry weight (cv. Hunisa and Domina) were used in gene expression experiments (Table S1). The ovaries at pre-anthesis and berries at *véraison* from each genotype were phenotyped as described by Houel et al (submitted). The average carpel number, the average berry flesh weight at *véraison*, the average berry width and length at *véraison*, the average seed weight and average number of seeds per berry were used in the association tests.

### **DNA extraction**

For sequencing experiments, DNA was extracted according to Houel et al (2010). For genotyping experiments, total DNA was extracted from 50 mg of young leaves using the DNeasy96 Plant Kit (Qiagen) according to the manufacturer and was quantified using the Quant-iT™ PicoGreen® dsDNA Assay Kit (Invitrogen).

### **RNA extraction and reverse transcription**

Flowers at stages 51, 53, 55, 57 and 61 (anthesis) according to BBCH international scale (Lorenz et al, 1994) and berries at a week and two week after anthesis) from four genotypes contrasted for berry size (Table S1) were collected during two years (2009 and 2010) and immediately frozen in liquid nitrogen. Total RNA was extracted from 50 mg of corresponding organs using the Spectrum™ Plant Total RNA Kit (Sigma) according to the manufacturer and was quantified using the Quant-iT™ RiboGreen® RNA Assay Kit (Invitrogen). RNA quality and the absence of DNA were checked by electrophoresis of 1 µg of total RNA for each sample on an agarose gel. Independent complementary DNA (cDNA) synthesis reactions were made on seven different flower and berry development stages for 4 different genotypes for two years. Reverse Transcription (RT) was performed on 1 µg of total RNA with oligodT primer (18 mer) and the Superscript II Rnase H reverse transcriptase (Invitrogen) according to

**Table 1.** Primers used for q-RT-PCR assays.*Candidate genes for berry size variation*

Gene id 12X Genome Assembly	Gene name	Primer name	Primer forward	Primer reverse	amplicon lenght on cDNA	PCR efficiency
GSVIVG01013435001	<i>VvHMGR1</i>	VVC2942J*	GAAGGATACGAGACACTGGA	TAACTCCATCCCAACTTCTG	257	0.99
		VVC2942K	TAAGGAGCATTATCCCAAC	TGTAACTCCATCCCAACTTC	105	0.15
		VVC2942L	CAGAGGTAGAGGAAGTGGAA	CTCCTTGTAGCTTTTGATGG	110	0.99 <sup>§</sup>
GSVIVG01025945001	<i>VvAGL11</i>	VVC21934G	CAAGCAAACATGGTATCAAC	TGAAAGAAATTGCGAGAAAC	68	0.93
		VVC21934H	CAAGCAAACATGGTATCAAC	TCATATTGGGCTGAAAGAAA	79	0.79
		VVC21934I*	CATCCAGGCATTAGTTTCTC	TTCTTATCAGGAAGTGGGTAG	84	0.99
GSVIVG01037533001	<i>VvJMJ</i>	VVC0413E*	GTGCTTTACATCAAAGAATGTG	TAACCCTGGAGTATCCCAAC	111	0.98
		VVC0413F	GCTTTACATCAAAGAATGTGG	AGTATCCCAACAGCAGCAA	100	0.82

*Reference genes*

Gene id 12X Genome Assembly	Gene name	Reference	Primer forward	Primer reverse	amplicon lenght on cDNA	PCR efficiency
GSVIVG01007521001	<i>GAPDH</i>	Reid et al, 2006	TTCTCGTTGAGGGCTATTCCA	CCACAGACTTCATCGGTGACA	70	0.99
GSVIVG01025146001	<i>EF1-α</i> **	Reid et al, 2006	GAAGTGGGTGCTTGATAGGC	AACCAAAATATCCGGAGTAAAAGA	150	0.99
GSVIVG01025191001	<i>SAND</i>	Reid et al, 2006	CAACATCCTTTACCCATTGACAGA	GCATTTGATCCACTTGCAGATAAG	76	0.97
GSVIVG01038617001	<i>UBQ-L40</i> **	Fernandez et al, 2007	AGTAGATGACTGGATTGGAGGT	GAGTATCAAAACAAAAGCATCG	174	0.99
		Reid et al, 2006	CATAACATTGCGGCAGATCA	TGGTGGTATTATTGAGCCATCCTT	80	0.99
GSVIVG01026580001	<i>Actin</i>	Reid et al, 2006	CTTGCATCCCTCAGCACCTT	TCCTGTGGACAATGGATGGA	82	0.99
GSVIVG01028330001	<i>MDH</i>	Reid et al, 2006	CCATGCATCATCACCCACAA	GTCAACCATGCTACTGTCAAAACC	72	0.99
GSVIVG01025462001	<i>α-Tubulin</i>	Reid et al, 2006	CAGCCAGATCTTCACGAGCTT	GTTCTCGCGCATTGACCATA	119	0.99
GSVIVG01016276001	<i>Aquaporin</i> **	Reid et al, 2006	TCCGCCAAGGACTATCATGAC	CGCAATCAGAGCCCTGTAGAA	90	0.99

\* Primer pairs selected for candidate genes expression study and

\*\* Reference genes with a constance expression during flower and berry developement selected

<sup>§</sup> problem on dissociation curve

the manufacturer's instruction. Negative controls on the same samples without Superscript II Rnase H reverse transcriptase (RT-) were also realized.

### Quantitative Real-Time PCR

The primers used for the quantitative Real-Time PCR Polymerase Chain Reaction assays (q-RT-PCR) are described in Table 1. Three primers per candidate gene were designed on one exon or as much as possible on two exons flanking an intron using the Primer3 software v.0.4.0 (Rozen et al, 2000) with the following parameters: optimal length 21 bases, hybridization temperature between 53 and 57 with an optimum at 55°C, amplicon length 60-150 bp on cDNA. Nine reference genes known to be constitutively expressed during flower and berry development were also selected (Fernandez et al 2007; Reid et al 2006; Table 1). The primer pairs were first tested on a range of dilutions (1/20, 1/100, 1/1000, 1/5000) made from a pool of cDNA from all samples in order to generate a standard curve and assess their PCR efficiency. Quantitative Real-Time PCR (q-RT-PCR) experiments were performed in three replicates (technical repeats) from samples collected in 2009 and 2010 (biological repeats) for each primer pair and included three controls containing no template (RT-). Q-RT-PCR reactions were performed in 15 µl, with 2.5 µl of each dilution, 2.5 µl of a mix primer pair at 1µM, and 7.5 µl SYBRGreen PCR master mix 2X (Eurogentec). All reactions were performed with the ABI PRISM 7900 HT sequence Detection System (Applied Biosystem) as follows: 95°C, 10 min; 40 cycles of 95°C 15 s and 60°C 1 min and a dissociation step to discriminate primer dimers from the PCR product. Using the SDS software provided by the manufacturer, the optimal cycle threshold ( $C_t$ ) was determined from the dilution series, with the raw expression data derived. Band size and the absence of other non-specific products were checked by agarose gel electrophoresis of the PCR products. For each candidate gene, the primer pair with the best PCR efficiency (between 97% and 100%) and specific to the corresponding gene was chosen (Table 1). For reference genes the same criteria were used and the constitutively expressed of each gene during flower and berry development was checked. Finally, two reference genes were used: *EF1-α* and *Aquaporine* (Table 1). Q-RT-PCR were performed on the dilution series from the mixed cDNA solution (1/20, 1/100, 1/1000) and on a 1/100 dilution of each sample as indicated previously. The data obtained from the two constitutively expressed genes were used to normalize the data of expression of each candidate gene. Correct size of amplicon and the absence of multiple PCR products were checked by electrophoresis of 5 µl of each PCR product on an agarose gel. The rest of each





PCR product was sequenced and sequences were blast on 12X version of genome annotation to check the amplification of corresponding gene.

### **Gene fragments amplification and sequencing**

Based on the genome annotation provided by Jaillon et al (2007), 90 primer pairs from candidate gene for berry size were designed according to the procedure described by Houel et al 2010 in order to amplify the maximum of exonic sequences from candidate genes for berry size variation (Table S2). All PCR amplifications on DNA from the 26 genotypes (Table S1) were carried out as described by Philippe et al (2010). The amplicon sequences were then aligned on the last 12X version of the genome sequence ([http://urgi.versailles.inra.fr/cgi-bin/gbrowse/vitis\\_12x\\_pub/](http://urgi.versailles.inra.fr/cgi-bin/gbrowse/vitis_12x_pub/)) and two of them (VVC7282A and VVC7282B) did not correspond to a gene model anymore.

### **Sequence data analysis and SNP choice**

Raw data were aligned and trimmed using either the Genalys v.2.8.3b software for Macintosh (Takahashi et al, 2003) or the Staden software v.2.0.0 (Staden, 1996). SNP were manually curated and insertion / deletion (INDEL) added when needed. A maximum of SNP or INDEL (<7 pb) both with a frequency > 5%, were chosen for the candidate genes (Table S3) and one SNP per gene with a frequency > 5% was chosen for each of the 77 genes randomly spread over the genome described in Houel et al 2010 (Table S4). These choices were done according to the SNPLEX™ design rules (Applied Biosystems Assays-by-Design Service SM; <http://www.appliedbiosystems.com>). The sequence of 70 bp flanking the SNP or INDEL on both side, were blasted against the grapevine genome 12X sequence, in order to discard SNP with multilocus flanking sequences. Three SNPLEX panels of 48 SNP probes were finally designed: two panels for candidate gene for berry size and one panel for random genes.

### **SNP and INDEL genotyping**

SNPLEX genotyping was carried out on DNA at a final concentration of 2-10 ng in a final volume of 4µl. Three sets of 48 SNP probes were analysed according to the manufacturer's protocol (<http://www.appliedbiosystems.com>). Samples were run on the 3730xl DNA Analyzer (Applied Biosystems) and data were analysed and manually edited when needed, using Gene Mapper v.3.7 software (Applied Biosystems).

**Table 2.** List of candidate genes for berry size variation.

<i>Vitis</i> gene name <sup>§</sup>	<i>Vitis</i> functional annotation	Criteria <sup>§</sup>	Putative orthologous gene (Species)*	Putative orthologous gene function	Reference
<i>VvAGL11</i>	Transcription factor	Co-localisation with berry weight & seed number QTL	<i>Agamous like 11</i> (Arabidopsis)	MADS box expressed in the carpel	Mejia et al, 2011; Boss et al, 2002; Diaz-Riquelme et al, 2009
<i>VvPMEI2</i>	Pectin Methylesterase Inhibitor	Co-localisation with berry weight QTL	<i>MMI9.1</i> (Arabidopsis)	Pectinesterase	N. Mejia unpublished result; Barnavon et al, 2001
<i>VvSPY</i>	Binding protein	Co-localisation with berry weight QTL	<i>SPINDLY</i> (Arabidopsis)	Repressor of GA and positive regulation of cytokinin signalling	N. Mejia unpublished result; Silverstone et al, 2007
<i>VvRPL24</i>	structural constituent of ribosome	Co-localisation with berry weight QTL	<i>Ribosomal protein like 24</i> (Arabidopsis)	Auxin mediated signaling pathway and gynoeceium development	N. Mejia unpublished result; Nishimura et al, 2005
<i>VvFw2.2.2</i>	Unknown	Bibliography	<i>Fw2.2</i> (Tomato)	Negative regulator of cell division in fruit development	Frary et al, 2000
<i>VvFw2.2.1</i>	Unknown				
<i>VvFw2.2.3</i>	Unknown				
<i>VvSUN</i>	Calmodulin binding protein	Bibliography	<i>SUN</i> (Tomato)	Control of fruit shape	Xiao et al, 2008
<i>VvOvate1</i>	Unknown	Bibliography	<i>Ovate</i> (Tomato)	Control the transition from round to pear-shaped fruit	Liu et al, 2002
<i>VvOvate2</i>	Unknown				
<i>VvHMGR1</i>	HMGCoA reductase	Bibliography	<i>HMG1</i> (Arabidopsis)	Involved in cell elongation in Arabidopsis	Suzuki et al, 2004
<i>VvHMGR2</i>	HMGCoA reductase				
<i>VvHMGR3</i>	HMGCoA reductase				
<i>VvAP3</i>	Transcription factor	Bibliography	<i>Apetala 3</i> (Arabidopsis)	MADS box involved in flower development	Boss et al, 2002; Poulin et al, 2007
<i>VvAGL22.2</i>	Transcription factor	Bibliography	<i>Agamous like 22</i> (Arabidopsis) <sup>§</sup>	MADS box involved in flower development	L. Torregrosa unpublished result
<i>VvAGL17.1</i>	Transcription factor	Bibliography	<i>Agamous like 17</i> (Arabidopsis)	MADS box	Diaz-Riquelme et al, 2009
<i>VvAGL12</i>	Transcription factor	Bibliography	<i>Agamous like 12</i> (Arabidopsis)	MADS box involved in flowering time	L. Torregrosa unpublished result
<i>VvAGL8</i>	Transcription factor	Bibliography	<i>Agamous like 8 - Fruitfull</i> (Arabidopsis)	MADS box involved in flower and fruit development	Gu et al, 1998; Diaz-Riquelme et al, 2009
<i>VvMADS1</i>	Transcription factor	Bibliography	<i>Agamous</i> (Arabidopsis)	MADS box involved in fruit development	Boss et al, 2001
<i>VvMADS6</i>	Transcription factor	Bibliography	<i>Apetala 1</i> (Arabidopsis) <sup>§</sup>	MADS box involved in flower development	L. Torregrosa unpublished result
<i>VvYABBY2</i>	YABBY protein	Bibliography - SSH (WT)	<i>Fasciated</i> (Tomato)	Controls carpel number during flower and/or fruit development	Fernandez et al, 2007; Cong et al, 2008
<i>VvMADS9</i>	Transcription factor	SSH (M)	<i>Pistilata</i> (Arabidopsis)	MADS box involved in flower development	Fernandez et al, 2007; Poulin et al, 2007
<i>VvJMJ</i>	Jumonji domain protein	SSH (WT)	<i>F2K11.14</i>	Transcription factor jumonji	Fernandez et al, 2007
<i>VvMADS4</i>	Transcription factor	SSH (WT)	<i>Agamous like 9 - Sepallata 3</i> (Arabidopsis)	MADS box involved in flower development	Fernandez et al, 2007; Diaz-Riquelme et al, 2009
<i>VvHB13</i>	Transcription factor	SSH (WT) - Microarrays (WT)	<i>Homeo Box 13</i> (Arabidopsis)	Homeotic protein	Fernandez et al, 2007
<i>VvSVP2</i>	Specific tissue protein	SSH (WT) - Microarrays (WT)			Fernandez et al, 2007
<i>VvBURP1</i>	BURP domain protein	SSH (WT) - Microarrays (WT)	<i>Unknown seed protein like 1</i> (Arabidopsis)	Seed development protein	Fernandez et al, 2007
<i>VvHD2C</i>	Zinc finger protein	Microarrays (M)	<i>HD2C</i>	Histone deacetylase 2C	Fernandez et al, 2007

<sup>§</sup> Candidate gene co-localized with a QTL for berry size or for seed number; Bibliography: genes involved in fruit size variation and/or flower development in other fruits or expressed in flower and/or berry in grapevine; Gene expressed in Ugni Blanc wild-type and not in Ugni Blanc fleshless berry mutant (WT) or the reverse (M) based on Suppression Subtractive Hybridization (SSH) or microarray experiments (M).

\* The orthologous gene in *Vitis* was chosen by the reciprocal best hit method.

<sup>§</sup> The *V. vinifera* orthologous gene corresponds to the second hit in the reciprocal blast.

<sup>¶</sup> The gene id provided by Jaillon et al (2007) is in Table S2.

### Association genetics

The population structure was calculated using the Structure software (Pritchard et al, 2000) based on the genotypes at 20 SSR markers well spread along the 19 chromosomes (Lauco et al, 2011; R. Bacilieri unpublished results). A kinship matrix was generated based on the genotypes at the 45 SNP from the random genes spread over the genome (Table S4). A General Linear Model (GLM) test using the structure of the sample and a Mixed Linear Model (MLM) test using both the structure of the sample and the kinship matrix were carried out to detect possible associations between the SNP and INDEL from the candidate genes and ovary and berry traits variation using the TASSEL software v.2.1 ([http://www.maizogenetics.net/index.php?option=com\\_content&task=view&id=89&Itemid=119](http://www.maizogenetics.net/index.php?option=com_content&task=view&id=89&Itemid=119)). A Bonferroni correction (P-value <0.05) was applied to control false-positives.

## Results

### Choice of the candidate genes

Twenty-eight candidate genes for berry size variation were chosen based on different criteria and are listed in Table 2. Four genes were chosen based on their presence within a QTL regions involved in berry size variation. *VvAGLII* maps to a QTL interval explaining 27% to 78% of phenotypic variation for seed weight and 21% to 44% of berry weight variation (Doligez et al 2002; Cabezas et al, 2006; Mejia et al, 2007; Costantini et al, 2008; Mejia et al, 2011). It was also shown to have a carpel-specific RNA expression in grapevine (Boss et al, 2002) and is putatively orthologous to *AGLII* involved in ovule identity in Arabidopsis (Favaro et al, 2003) and to two MADS-box genes from petunia (*FBP7* and *FBP11*) which were involved in ovule and seed development (Columbo et al, 1995; Angenent et al, 1995; Columbo et al, 1997). *VvPMEI2* co-localized together with four other genes with a QTL explaining 17% of the berry weight variation in a cross between two seedless varieties (N. Mejia, personal communication). It encodes a putative pectin methylesterase inhibitor that might be involved in the control of berry cell wall expansion during berry development and was found expressed in berries (Barnavon et al, 2001). *VvSPY* and *VvRPL24* co-localized together with 98 other genes with a QTL explaining 13% of the phenotypic variation of berry weight (N. Mejia, personal communication). The putative orthologous gene to *VvSPY*, *SPINDLY* regulates gibberellin and cytokinin signaling pathways in Arabidopsis (Jacobsen et al, 1993; Silverstone et al, 2007) and is expressed in grapevine berry (Terrier et al, 2005).



while *VvRPL24* encodes a ribosomal protein like that are involve in auxin mediated signalling pathway and gynoecium development in Arabidopsis (Nishimura et al, 2005).

Seventeen other genes were chosen because they were putative orthologs (by the Bi-Directional Best Hit (BDBH) method) to genes that were involved into fruit size variation in tomato or in fruit or flower development in Arabidopsis or/and were expressed in flower and/or berry in grapevine. *Fw2.2* is a negative regulator of cell division in tomato fruit (Frary et al, 2000) and *Fasciated* is involved in the tomato carpel number variation (Cong et al, 2008) both involved in fruit size variation. In grapevine, three genes are homologous to *Fw2.2* and were named *VvFw2.2.1*, *VvFw2.2.2*, *VvFw2.2.3* and one to *Fasciated*, called *VvYABBY2*. Three other genes *VvOvate1*, *VvOvate2* and *VvSUN* were respectively putatively orthologous to *Ovate* and *SUN* involved in tomato fruit shape (Liu et al, 2002; Xiao et al, 2008). The grapevine 3-hydroxy-3-methyl-glutaryl-CoA reductase family (*VvHMGR1*, *VvHMGR2*, *VvHMGR3*) was studied because all the family had similar sequences to *HMGR1*, which is involved in cell elongation in Arabidopsis (Suzukiet al, 2004) and because the HMGR family has been showed to be expressed in early stages of fruit development in tomato (Narita and Gruissem, 1989) and melon (Kobayashi et al, 2002), when rapid cell divisions are occurring. Moreover, the inhibition of their activity reduced the growth of fruit in tomato (Narita and Gruissem, 1989) and in avocado (Cowan et al, 1997).

Seven putative grapevine MADS box genes (*VvAP3*, *VvAGL22.2*, *VvAGL17.1*, *VvAGL12*, *VvAGL8*, *VvMADS1* and *VvMADS6*) were also selected based on the fact that their putative orthologs in Arabidopsis are involved in flower or fruit development (Gu et al, 1998; Jack, 2001) and/or that they were expressed in grapevine flowers and/or berries (Boss et al, 2002; Sreekantan et al, 2006, Fernandez et al, 2007; Chatelet et al, 2007; Poulin et al, 2007; Diaz-Riquelme et al, 2009; L. Torregrosa unpublished results).

Finally, seven other genes were chosen (*VvMADS9*, *VvJMJ*, *VvMADS4*, *VvHB13*, *VvSVP2*, *VvBURP1*, *VvHD2C*) based on their differential expression in young berries between Ugni Blanc and its fleshless mutant (Fernandez et al, 2007) either in Suppression Subtractive Hybridization (SSH) or microarray experiments (Fernandez et al, 2007).

### SNP discovery and genotyping on candidate genes for berry size

One to eight primer pairs were designed to cover the complete coding sequence of each of the 28 candidate genes and the set of 80 useful primers pairs that was used for sequencing is described in Table S2. Thirty-two of them were used for sequencing in forward, seven in

**Table 3.** SNP significantly associated with the variation of berry trait at maturity.

Gene id	Gene name	SNP type	SNP position	SNP position on the genome sequence	F test	P-value <sup>§</sup>	Trait
GSVIVG01019761001	<i>VvJMJ</i>	[T/A]	Exon*	3117781	9.6	9.52E-05	Berry length
					8.6	2.51E-05	Berry weight
GSVIVG01013435001	<i>VvHMGR1</i>	[T/C]	Exon*	497314	9.1	1.56E-05	Berry length
		[G/A]	Intron	495082	8.9	1.87E-05	Berry length
GSVIVG01025945001	<i>VvAGL11</i>	[C/T]	Intron	26889528	14.8	9.22E-07	Seed number
		[G/T]	Intron	26889822	13.4	3.07E-06	Seed number

\* non synonymous mutation

§ Significant P-value after Bonferroni correction at level 0.05

reverse and 41 on both strands in the 26 genotypes maximizing the genetic diversity of cultivated grapevine, totalling in average 58,620 bp of sequence per genotype. In average, 23 individual sequences were obtained per candidate gene. The analysis of the whole set of sequences resulted in the detection of a total of 704 SNP or insertion/deletion events.

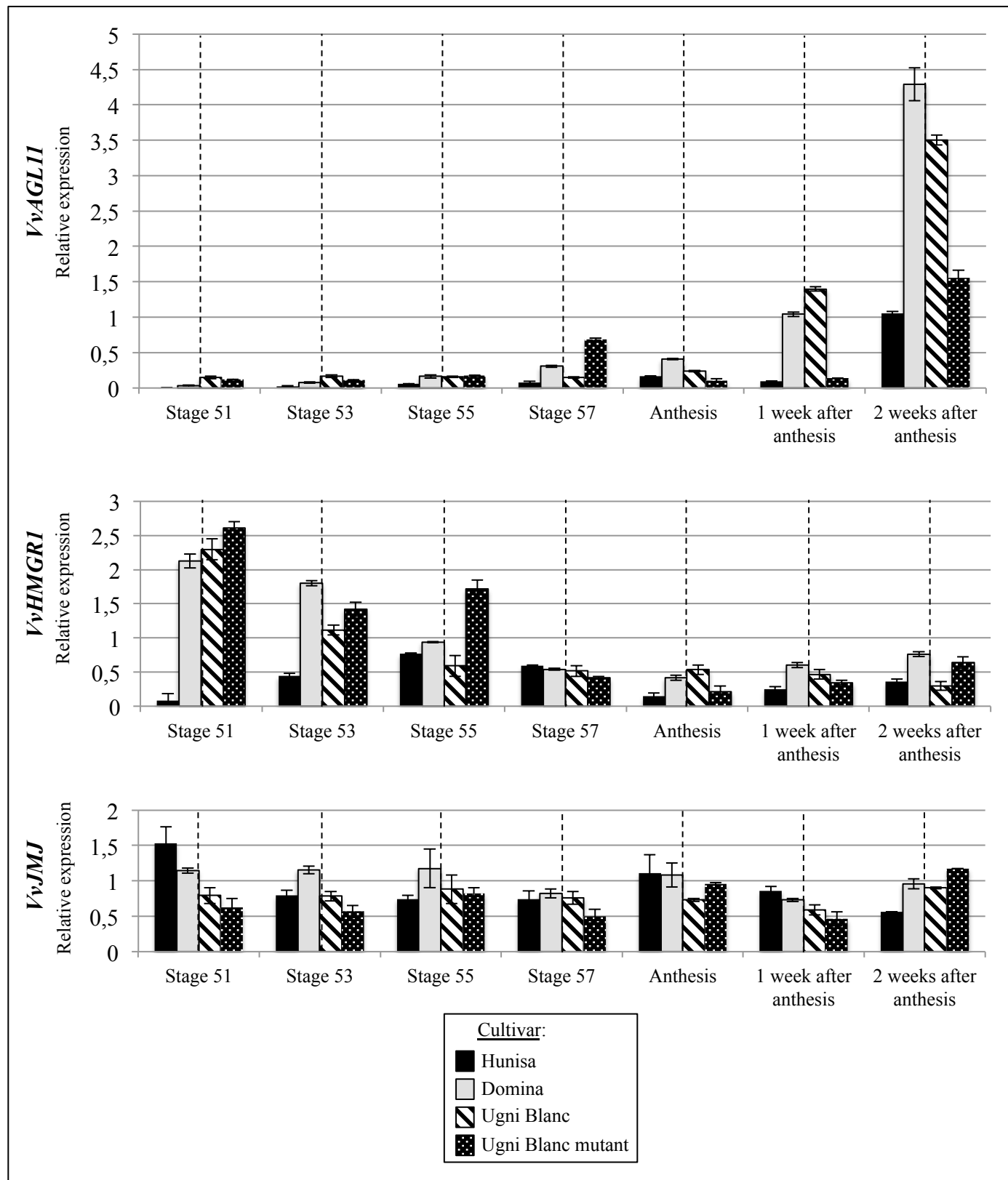
Taking into account the criteria for the SNPLex design, only 248/704 of these polymorphisms could be kept and this number was again reduced to 145 in 26 out of the 28 candidate genes after filtering for specificity. No useful SNP or insertion/deletion in *VvFw2.2.2* and *VvFw2.2.3* could meet the criteria for SNPLex genotyping. Finally, 96 of these polymorphisms targeting the 26 remaining genes were compatible for multiplex processing and arranged in two SNPLex panels (Table S3). Another SNPLex panel of 48 polymorphisms was design using the same criteria from 48 genes out of 77 genes randomly spread over the grapevine and re-sequenced on the same sample of individuals in Houel et al, 2010 (Table S4).

The genotyping of the 240 cultivated grapevine genotypes resulted in 83 loci with useful results on the candidate genes and 45 on the random genes. Most of the discarded loci corresponded to primers that still appeared to target multiple loci (14 out of 16 cases; Tables S3 and S4). Finally, the 26 candidate genes were represented by one locus at least, whereas the final set of random gene genotyped was reduced to 45 at this stage of the analysis (Table S4). The genotyping data from the set of random gene was used to calculate a kinship matrix.

### Association genetics

The association tests were performed following either the General Linear Model (GLM) test using the structure of the sample or a Mixed Linear Model (MLM) test using both the structure of the sample and the kinship matrix. A Bonferroni correction at level 0.05 was applied for multiple hypotheses testing. MLM provided six significant associations, while GLM appeared slightly less conservative, detecting one additional significant association (data not showed).

The six significant associations obtained with the MLM method are presented in Table 3. One SNP in the *VvJMJ* gene encoding a putative transcription factor was associated with berry length and weight variations at *véraison*. This SNP corresponded to a non-synonymous mutation located in PLU-1 predicted domain of the protein. This domain may have a role in DNA-binding and transcription and is closely associated with the malignant phenotype of breast cancer in human (Yamane et al, 2007). Two SNP from the *VvHMGR1* gene encoding a putative 3-hydroxy-3-methyl-glutaryl-CoA reductase were found associated with berry length



**Figure 1.** Expression patterns obtained by q-RT-PCR for *VvJMJ*, *VvHMGRI*, *VvAGL11* in four genotypes contrasted for berry size during flower and early berry development. Stage 51, 53, 55, 57 range from the onset of inflorescence to a flower entirely developed according to BBCH international scale (Lorenz et al, 1994). Expression values have been normalized with the *EF1-α* and *Aquaporin* expression data. Bars correspond to the standard deviation of the PCR triplicates and the two biological replicates.



at *véraison* but interestingly not with berry weight, suggesting that the two genes are involved into different mechanisms. One of the two SNP was a non-synonymous mutation located in the tetramerization domain of the protein. Finally, two SNP in introns of the *VvAGL11* gene were found associated only with seed number.

### Candidate genes expression during flower and early berry development

Pattern of expression of the three genes *VvJMJ*, *VvAGL11* and *VvHMGR1* during flower and early berry development were carried out in order to confirm their role in berry trait variations. The candidate gene expression was performed by q-RT-PCR assay in flowers from five stage of development corresponding to onset of inflorescence to anthesis, and in berries at one and two weeks after anthesis. Flowers and berries at different stages of development were sampled from four genotypes that were selected for their contrasted berry phenotypes: *cv.* Hunisa and Domina which produce berries weighting at maturity respectively in average 6 g and 1 g, *cv.* Ugni Blanc which produces berries of 2 g and its somatic mutant which produces fleshless berries weighting in average 0.3 g. Two biological replicates corresponding to samples collected in 2009 and in 2010 for each stage and genotype were analysed. When these two replicates showed the same variation between the two years, the results were pooled. The average expression patterns of the candidate gene during the two years are presented in Figure 1.

All the candidate genes were expressed during flower and berry development. The pattern of expression of *VvJMJ* was not different between genotypes contrasted for berry size, nor between stages of flower and berry development. *VvHMGR1* and *VvAGL11* had opposite expression patterns. *VvHMGR1* decreased from the onset of flower development (stage 51 and 53; Figure 1) to berry stages whereas *VvAGL11* showed a maximum of expression in berry two weeks after anthesis (Figure 1) and low expression in all flower stages. The cultivar with large and oval berries, Hunisa, showed a clearly different pattern of expression from the other cultivars, all carrying small and round berries, at the stages where *VvHMGR1* and *VvAGL11* peaked, with a significantly lower expression in both cases. Finally, Ugni Blanc mutant retained a high expression of *VvHMGR1* longer than the other small berry cultivars (until stage 55) and its pattern of expression of *VvAGL11* was strangely similar to the pattern of expression in Hunisa, with a peak occurring later (Figure 1).

## Discussion

Eighty-three SNP identified in twenty-six candidate genes were tested for their possible association with six ovary or berry traits. Recent studies have shown that the relatedness between genotypes (kinship) performed better control of false positives than just overall population covariance (structure) (Zhao et al, 2007). Here we tested both models and found not much difference between them. This can be explained either by the kinship effects being weak compared to the strength of the associations detected or by the fact that 45 SNP from 45 random genes spread over the genome were not sufficient for in depth kinship estimation. The first hypothesis seems to be the best because the sample used in the present study was chosen to minimize the familial relatedness between genotypes. We chose to apply a rather conservative Bonferroni correction for each test to minimize false positives, knowing that genes with minor effects would not be detected. In parallel, taking into account both the structure of the sample and relatedness of individuals as variables in the model can increase the occurrence of false negatives. This has been observed for flowering time in maize which variation is linked to the structure of the population (Flint-Garcia et al, 2005; Camus-Kulandaivelu et al, 2006). In grapevine, there are indeed morphological differences between cultivars of Eastern and Western Europe: Eastern Europe cultivars are mostly composed of table grape cultivars with large berries whereas Western Europe cultivars mainly comprise wine grapes with small berries (Negrul, 1938; Levadoux, 1956). However, we showed in a recent study (Houel et al, submitted) that berry size variation in cultivated grapevine was not superimposed with the evolutionary history that led to the genetic structure in Groups WW, WE and TE.

Six associations involving five SNP in three genes, *VvAGL11*, *VvJMJ* and *VvHMGR1*, out of the 26 screened were found. The analysis of the expression of the corresponding genes in flowers and berries at development stages chosen within the period of the development shown to be involved in berry size variation by Houel et al (submitted) was then carried out. The three genes were indeed expressed in the berries, as expected from the function of their putative homologs in other plants or former experiments in grapevine. None of them were the putative orthologs of the genes currently identified for their role in variation of tomato size, excepted for the *VvHMGR1* which belongs to a gene family showed to be involved in tomato fruit growth (Narita and Gruissem, 1989). It is possible that our experimental design was not sensitive enough to evidence the role of the other candidate genes but it may also illustrate

differences between the two plant families in the control of fruit size. Moreover, *fw2.2* has been shown to be responsible for the fruit size change between the wild and the cultivated tomatoes. It is possible that some of the genes involved in berry size variation in grapevine will be only evidenced through the comparison between the wild and the cultivated compartments. Finally, false negative may be due to the low number of SNP tested per gene (from one to eight only) due to SNPlex™ design rules and the fact that none of them was in promoter regions associated to local high recombination frequency. Indeed, we could not get reliable results of promoter re-sequencing in the SNP discovery panel due to lots of heterozygous INDEL making unreadable the Sanger sequences.

*VvAGL11* encodes a putative Agamous like 11 protein well known for its role in the control of ovule and seed identity in Arabidopsis and Petunia (Rounsley et al, 1995; Columbo et al, 1995; Angenent et al, 1995; Columbo et al, 1997). In grapevine, the expression pattern of this gene was consistent with its putative function: it was lowly expressed during flower development and showed a peak of expression after anthesis, once the seeds start to develop after fertilization (also seen in Mejia et al, 2011). This gene co-localized both with a QTL explaining up to 44% of the variation of berry size and with the major QTL for seedlessness (Doligez et al, 2002; Mejia et al, 2007; Constantini et al, 2008; Cabezas et al, 2006; Mejia et al, 2011). Two hypothesis are yet still unresolved to explain the co-occurrence of these two QTL: (i) a pleiotropic effect of the gene underlying the two traits or (ii) the existence of two tightly linked QTL in this region containing only 8 genes (Mejia et al, 2011). The first hypothesis seems more likely, since a positive intra-genotypic relation was found between the seed number and the berry size by several authors (Doligez et al, 2002; Mejia et al, 2007; Constantini et al, 2008; Cabezas et al, 2006; Houel et al, submitted). In this study, two SNP in *VvAGL11* were found associated with the seed number per berry but not with the berry weight, which was consistent with the fact that no inter-genotypic correlation was found between the seed content and the berry weight in the same set of genotypes (Houel et al, submitted). This could support the second hypothesis. However, *VvAGL11* also showed here different patterns of expression between genotypes contrasted for average berry weight and seed weight (respectively 6 g and 0.07 g for Hunisa at maturity and 1 g and 0.03 g for Domina) while these varieties contained in average two seeds per berries. In Mejia et al (2011) sequence variation in *VvAGL11* was found associated with seed and berry weight. Three factors could explain the differences with the present study. First, here the association between the SNP and the seed weight was just below the Bonferroni threshold, which is known to be a rather

conservative test. Second, the association study performed in Mejia et al (2011) was performed in a panel of genotypes segregating for seedlessness and third Mejia et al (2011) showed that polymorphisms in different parts of the genes (promoter and the last exons) had different effects on phenotypic variation. Interestingly, the cultivar Ugni Blanc mutant producing round shape seeds showed for this gene a different pattern of expression to the Ugni Blanc wild type with produced elongated shape seeds (Fernandez et al, 2006a). The role of this gene on seed number, weight and shape but also on berry weight remains therefore to be elucidated.

Two SNP from *VvHMGR1*, a putative 3-hydroxy-3-methyl-glutaryl-CoA reductase (HMGR) were found associated with berry length. The HMGR catalyses the conversion of 3-hydroxy-3-methylglutaryl coenzyme A into mevalonic acid in the isoprenoid biosynthesis pathway (Chappell 1995, McGarvey and Croteau, 1995). Isoprenoid compounds are, among other things, the precursors of plant growth regulators like cytokinine, abscisic acid and gibberellins (Enjuto et al, 1994). Several studies have indicated that the HMGR activity during early fruit development in tomato, avocado and melon is associated with cell division and fruit growth (Narita and Gruissem 1989; Cowan et al, 1995; Jelesko et al, 1999; Kobayashi et al, 2002), while it would be involved in cell elongation in Arabidopsis (Suzuki et al, 2004). Moreover, in avocado small fruits showed a lower activity of the HMGR correlated with an increase of the content in abscisic acid compared to large fruits (Cowan et al, 1995). However none of the studies in fleshy fruits has analysed the expression or activity of HMGR during flower development. Yet, in the present study *VvHMGR1* gene was shown highly transcribed at the beginning of flower development, except for the cultivar Hunisa that was the only genotype producing large and oval berries. The transcription of the *VvHMGR1* gene was analysed only at the beginning of the berry development and it is therefore not possible to know if its expression would decrease later during the berry development like it was observed for other fruits, when this gene was associated with fruit size variation (Narita and Gruissem 1989; Cowan et al, 1995; Jelesko et al, 1999; Kobayashi et al, 2002). In parallel, treatments of berries from seedless cultivars (Danuta Blanc, Alvina Noir and Sultanine Blanc) with gibberellins increased berry length up to 20% and berry width only by 3% in average (C. Rennes and JM. Boursiquot, personal communication) suggesting a role of gibberellin on berry length. The precise role of *VvHMGR1* either in cell elongation or in directional cell multiplication and its relationship with *VvAGL11* in interaction with the gibberellins or with the synthesis of abscisic acid is still to be elucidated.

The same SNP in the *VvJMJ* gene was found to be involved with berry weight and berry length variation, suggesting a pleiotropic effect of this gene on fruit shape and fruit size, like it was observed in tomato for *Fasciated* and *Lc* (Tankley, 2004). A relation between these two traits was also suggested from phenotypic analysis in grapevine by Houel et al (submitted). This could be due to different molecular mechanisms controlling directions of cell divisions and/or cell expansion that could evolve more or less independently (Coombe, 1976; Tsukaya, 2003). *VvJMJ* encodes a putative transcription factor jumonji. The jumonji protein was found to be involved in transcriptional repression and/or chromatine regulation in human, Arabidopsis, yeast and fungus (Takeuchi et al, 2006; Tamaru, 2010). This gene was chosen at the beginning of this study on the basis of its differential expression in young berries between Ugni Blanc and its fleshless mutant in SSH experiments (Fernandez et al, 2007). The q-RT-PCR performed here did not however support these former results. Indeed the SSH cDNA library was constructed from three berry developmental stages (3, 6 and 15 day after anthesis or DAA; Fernandez et al 2007) while we sampled here only 6 and 15 DAA. It is possible that difference in expression between the two genotypes occurred earlier, at 3 DAA, when the intense cell divisions take place in the flesh tissues (Ojeda et al, 1999). We also showed in a previous study that the final berry shape seems not to be predictable at anthesis but later during berry development (Houel et al, submitted). This is in agreement with the fact that no difference of expression of *VvJMJ* was observed between cultivars with round berries (Domina, Ugni Blanc) and the cultivar with oval berries (Hunisa) during flower development and very early berry development. Expression patterning of the *VvJMJ* gene observed here did not bring any additional element to better understand its role in berry weight and length variation. However, the SNP found in the present study in this gene and in the *VvHMGR1* gene could be used for marker assisted breeding for berry weight and shape, independently of the seed number variation.

In conclusion, fruit size is an important trait in fruit crops. If the genetic and molecular control of fruit size variation has largely progressed in tomato, no gene involved in the regulation of fruit size variation in grapevine has been identified so far and the question about to which extent the knowledge accumulated in tomato may be transferable to grapevine was still open. Indeed, fruit size variation seems to be determined by common but also specific developmental mechanisms in different fleshy fruits (Houel et al submitted). Several studies have identified potential regions involved in berry and seed weight variation. This knowledge

was completed with the search for putatif orthologous genes identified in model plants, and an initial list of potential candidate genes can be derived but the validation and characterization of these genes is limited by biological properties of grapevine and the current state of genomic tools (no mutant library...). Here, we use association genetics as an additional screening of the involvement of these genes in the variation of fruit size, this method having previously been performed successfully in grapevine, and not requiring the creation of segregating population very long to implement in grapevine. Our study added supplementary evidences for the involvement of *VvAGL11* and *VvJMJ* respectively in seed and berry development. The general role of the *HMGR1* gene in berry development was also shown. The associations found between seeds and berry traits and the *VvAGL11* and *VvHMGR1* genes emphasized the interest of comparative functional genomics for the choice of candidate genes. The function of these genes will have to be clarified by reverse genetics, using strategies such as those proposed by Chaib et al (2010) for the accelerated characterization of traits controlling reproductive development. On the whole, such a screening by association genetics was useful to pinpoint the most interesting candidates.

## Aknowledgements

We thank Stéphane Nicolas and Thierry Lacombe for the design of plant material; Nilo Mejia for candidate genes into QTL for berry size; Isabelle Le Clainche and Rémi Bounon for technical assistance; Cécile Guichard for made-home script for SNP position on genome; Dominique Brunel for useful discussions; Jean-Pierre Péros for the set of reference gene sequences along the genome.

## Contributions of the authors

CH designed the experiments, chose the candidate genes, designed the primers for gene fragment re-sequencing, analyzed the results, collected flowers, berries and leafs, contributed to the RNA and DNA extraction, analyzed the results from q-RT-PCR experiments, realized association tests and drafted the manuscript. AC contributed to RNA extraction and with RB realized the q-RT-PCR experiments with some help of CH. LT provided part of the RNA samples of Ugni Blanc and Ugni Blanc mutant, contributed to the choice of the candidate genes. LLC helped for association tests and RB to the estimation on the structure of the sample. M-CL realized the sequencing of gene fragments and with AB and AC the hybridizations for the genotyping experiments. A-FA-B designed, supervised the experiments and corrected the manuscript.

## References

- Angenot GC, Franken J, Busscher M, Van Dijken A, Van Went JL, Dons HJM, Van Tunen AJ.** A novel class of MADS box genes is involved in ovule development in petunia. *Plant Cell* 1995, 17:569-1582
- Barnavon L, Doco T, Terrier N, Ageorges A, Romieu C, Pellerin P.** Involvement of pectin methyl-esterase during the ripening of grape berries: partial cDNA isolation, transcript expression and changes in the degree of methyl-esterification of cell wall pectins. *Phytochemistry* 2001, 58:693-701
- Boss PK, Sensi E, Hua C, Davies C, Thomas MR.** Cloning and characterization of grapevine (*Vitis vinifera* L.) MADS-box genes expressed during inflorescence and berry development. *Plant Sci* 2002, 162:887-895
- Boursiquot JM, Dessup M, Rennes C.** Distribution des principaux caractères phénologiques, agronomiques et technologiques chez *Vitis vinifera* L. *Vitis* 1995, 34:31-35
- Cabezas JA, Cervera MT, Ruiz-Garcia L, Carreno J, Martinez-Zapater JM.** A genetic analysis of seed and berry weight in grapevine. *Genome* 2006, 49:1572-1585

**Camus-Kulandaivelu L, Veyrieras JB, Madur D, Combes V, Fourmann M, Barraud S, Dubreuil P, Gouesnard B, Manicacci D, Charcosset A.** Maize adaptation to temperate climate: Relationship between population structure and polymorphism in the Dwarf8 Gene. *Genetics* 2006, 172: 2449-2463

**Chaïb J, Torregrosa L, Mackenzie D, Corena P, Bouquet A, Thomas MR.** The grape microvine - a model system for rapid forward and reverse genetics of grapevines. *Plant J* 2010, 62:1083-92

**Chappell J.** Biochemistry and molecular biology of the isoprenoid biosynthetic pathway in plants. *Annu Rev Plant Physiol and Plant Mol Biol* 1995, 46:521-547

**Chatelet P, Laucou V, Fernandez L, Sreekantan L, Lacombe T, Martinez-Zapater JM, Thomas MR, Torregrosa L.** Characterization of *Vitis vinifera* L. somatic variants exhibiting abnormal flower development patterns. *J Exp Bot*, 2007, 58:4107-18

**Cheniclet C, Rong WY, Causse M, Frangne N, Bolling L, Carde J-P, Renaudin J-P.** Cell expansion and endoreduplication show a large genetic variability in pericarp and contribute strongly to tomato fruit growth. *Plant Physiol* 2005, 139:1984-1994

**Colombo L, Franken J, Koetje E, Van Went J, Dons HKM, Angenent GC, Van Tunen AJ.** The *Petunia* MADS box gene FBP11 determines ovule identity. *Plant Cell* 1995, 7:1859-1868

**Colombo L, Franken J, Van der Krol AR, Wittich PE, Dons HKM, Angenent CG.** Down regulation of ovule-specific MADS box genes from *petunia* results in maternally controlled defects in seed development. *Plant Cell* 1997, 9: 703-715

**Cong B, Lui J, Tanksley SD.** Natural alleles at a tomato fruit size quantitative trait locus differ by heterochronic regulatory mutations. *Proc Acad Sci USA* 2002, 99:13606-13611.

**Cong B, Barrero LS, Tanksley SD.** Regulatory change in yabby-like transcription factor led to evolution of extreme fruit size during tomato domestication. *Nature Genetics* 2008, 40:800-804

**Coombe BG.** Relationship of growth and development to changes in sugars, auxins, and gibberellins in fruit of seeded and seedless varieties of *Vitis vinifera*. 1960. *Plant Physiol* 1960, 35:241-250.

**Coombe, BG.** The regulation of set and development of the grapeberry. *Acta Hort* 1973, 34: 261-273

**Costantini L, Battilana J, Lamaj F, Fanizza G, Grando MS.** Berry and phenology-related traits in grapevine (*Vitis vinifera* L.): from quantitative trait loci to underlying genes. *BMC Plant Biol* 2008, 8:38.

**Cowan AK, Moore-Gordon CS, Bertling I, Wolstenholme BN.** Metabolic Control of Avocado Fruit Growth (Isoprenoid Growth Regulators and the Reaction Catalyzed by 3-Hydroxy-3-Methylglutaryl Coenzyme A Reductase). *Plant Physiol* 1997, 114:511-518



**Díaz-Riquelme J, Lijavetzky D, Martínez-Zapater JM, Carmona MJ.** Genome-wide analysis of MIKCC-type MADS box genes in grapevine. *Plant Physiol* 2009, 149:354-69

**Doligez A, Bouquet A, Danglot Y, Lahogue F, Riaz S, Meredith P, Edwards J, This P.** Genetic mapping of grapevine (*Vitis vinifera* L.) applied to the detection of QTL for seedlessness and berry weight. *Theor Appl Genet* 2002, 105:780-795

**Emanuelli F, Battilana J, Costantini L, Le Cunff L, Boursiquot JM, This P, Grando MS.** A candidate gene association study on muscat flavor in grapevine (*Vitis vinifera* L.). *BMC Plant Biol* 2010, 9:10:241

**Enjuto M, Balcells L, Campos N, Caelles C, Arro M, Boronat A.** Arabidopsis thaliana contains two differentially expressed 3-hydroxy-3-methylglutaryl-CoA reductase genes, which encode microsomal forms of the enzyme. *Proc Natl Acad Sci USA* 1994, 91:927-931

**Fanizza G, Lamaj F, Costantini L, Chaabane R, Grando MS.** QTL analysis for fruit yield components in table grapes (*Vitis vinifera*). 2005. *Theor Appl Genet* 2005, 111:658-664

**Gu Q, Ferrándiz C, Yanofsky MF, Martienssen R.** The FRUITFULL MADS-box gene mediates cell differentiation during Arabidopsis fruit development. *Development* 1998, 125:1509-17

**Favaro R, Pinyopich A, Battaglia R, Kooiker M, Borghi L, Ditta G, Yanofsky MF, Kater MM, Colombo L.** MADS-box protein complexes control carpel and ovule development in Arabidopsis. *Plant Cell* 2003, 15:2603-11

**Fernandez L, Romieu C, Moing A, Bouquet A, Maucourt M, Thomas MR, Torregrosa L.** The grapevine fleshless berry mutation. A unique genotype to investigate differences between fleshy and non fleshy fruit. *Plant Physiol* 2006a, 140:537-547

**Fernandez L, Doligez A, Lopez G, Thomas MR, Bouquet A, Torregrosa L.** Somatic chimerism, genetic inheritance, and mapping of the fleshless berry (flb) mutation in grapevine (*Vitis vinifera* L.). *Genome* 2006b, 49:721-8

Fernandez L, Torregrosa L, Terrier N, Sreekantan L, Grimplet J, Davies C, Thomas M.R, Romieu C, Ageorges A. Identification of genes associated with flesh morphogenesis during grapevine fruit development. *Plant Mol Biol* 2007, 63:307-323

**Flint-Garcia SA, Thuillet AC, Yu JM, Pressoir G, Romero SM, Mitchell SE, Doebley J, Kresovich S, Goodman MM, Buckler ES.** Maize association population: a high-resolution platform for quantitative trait locus dissection. *Plant J* 2005, 44 1054-1064

**Fournier-Level A, Le Cunff L, Gomez C, Doligez A, Ageorges A, Roux C, Bertrand Y, Souquet JM, Cheynier V, This P.** Quantitative genetic bases of anthocyanin variation in grape (*Vitis vinifera* L. ssp. *sativa*) berry: a quantitative trait locus to quantitative trait nucleotide integrated study. *Genetics* 2009, 183:1127-39

**Frary, A, Nesbitt TC, Grandillo S, Knaap E, Cong B, Liu J, Meller J, Elber R, Alpert,KB, Tanksley SD.** *Fw2.2*: a quantitative trait locus key to the evolution of tomato

fruit size. *Science* 2000, 289:85-88

**Houel C, Bounon R, Chaïb J, Guichard C, Péros JP, Bacilieri R, Dereeper A, Canaguier A, Lacombe T, N'Diaye A, Le Paslier MC, Vernerey MS, Coriton O, Brunel D, This P, Torregrosa L, Adam-Blondon AF.** Patterns of sequence polymorphism in the fleshless berry locus in cultivated and wild *Vitis vinifera* accessions. *BMC Plant Biol* 2010, 10:284

**Houel C, Martin-Magniette M-L, Nicolas SD, Lacombe T, Le Cunff L, Franck D, Torregrosa L, Conéjéro G, Lalet S, This P, Adam-Blondon A-F.** Genetic variability of the berry size in grapevine (*Vitis vinifera* L.) involves ovary and early berry development associated factors. *J Exp Bot* submitted

**Jack T.** Relearning our ABCs: new twists on an old model. *TRENDS Plant Sci* 2001, 6:310-6

**Jacobsen SE, Olszewski NE.** Mutations at the SPINDLY locus of *Arabidopsis* alter gibberellin signal transduction. *Plant Cell* 1993, 5:887-896

**Jaillon O, Aury J-M, Noel B, Policriti A, Clepet C, Casagrande A, Choise N, Aubourg S, Vitulo N, Jubin C, Vezzi A, Legeai F, Huguency P, Dasilva C, Horner D, Mica E, Jublot D, Poulain J, Bruyere C, Billault A, Segurens B, Gouyvenoux M, Ugarte E, Cattonaro F, Anthouard V, Vico V, Del Fabbro C, Alaux M, Di Gaspero G, Dumas V, Felice N, Paillard S, Juman I, Moroldo M, Scalabrin S, Canaguier A, Le Clainche I, Malacrida G, Durand E, Pesole G, Laucou V, Chatelet P, Merdinoglu D, Delledonne M, Pezzotti M, Lecharny A, Scarpelli C, Artiguenave F, Pé E, Valle G, Morgante M, Caboche M, Adam-Blondon A-F, Weissenbach J, Quétier F, Wincker P.** The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 2007, 449:463-468

**Jelesko JG, Jenkins SM, Rodriguez-Concepcion M, Gruissem W.** Regulation of tomato HMGI during cell proliferation and growth. *Planta* 1999, 208:310-318

**Kobayashi T, Kato-Emori S, Tomita K, Ezura H.** Detection of 3-hydroxy-3-methylglutaryl-coenzyme A reductase protein Cm-HMGR during fruit development in melon (*Cucumis melo* L.). *Theor Appl Genet* 2002, 104:779-785

**Laucou V, Lacombe T, Dechesne F, Siret R, Bruno J-P, Dessup M, Dessup T, Ortigosa P, Parra P, Roux C, Santoni S, Vares D, Peros J-P, Boursiquot J-M, This P.** High throughput analysis of grape genetic diversity as a tool for germplasm collection management. *Theor Appl Genet* 2011, DOI 10.1007/s00122-010-1527-y

**Le Cunff L, Fournier-Level A, Laucou V, Vezzulli S, Lacombe T, Adam-Blondon A-F, Boursiquot J-M, This P.** Construction of nested core collections to optimize the exploitation of natural diversity in *Vitis vinifera* L. Subsp *sativa*. *BMC Plant Biology* 2008, 2:8-31

**Liu J, Van Eck J, Cong B, Tanksley SD.** A new class of regulatory genes underlying the cause of pear-shaped tomato fruit. *Proc Natl Acad Sci USA* 2002 99:13302-6

**Lorenz, D.H, Eichhorn K.W, Blei-holder H, Klose R, Meier H, Weber E.** Phänologische Entwicklungsstadien der Weinrebe (*Vitis vinifera* L. ssp. *vinifera*). *Vitic Enol Sci* 1994, 49:66-70

**McGarvey DJ, Croteau R.** Terpenoid metabolism. *Plant Cell* 1995, 7:1015-1026

**Mejía N, Gebauer M, Muñoz L, Hewstone N, Hinrichsen P.** Identification of QTL for seedlessness, berry size, and ripening date in a seedless × seedless table grape progeny. *Am J Eno Vitic* 2007, 58:499-507

*Mejia N, Soto B, Guerrero M, Casanueva X, Miccona M, Houel C, Le Cunff L, Boursiquot JM, Hinrichsen P, Adam-Blondon AF.* Molecular, genetic and transcriptional evidence for a role of *VvAGL11* in stenospermocarpic seedlessness in grapevine. *BMC plant biology*, 2011 in press

**Narita JO, and Gruissem W.** Tomato hydroxyl-methyl-glutaryl-CoA reductase is required early in fruit development but not during ripening. *Plant Cell* 1989, 1:181-190

**Nishimura T, Wada T, Yamamoto KT, Okada K.** The Arabidopsis STV1 protein, responsible for translation reinitiation, is required for auxin-mediated gynoecium patterning. *Plant Cell* 2005, 17:2940-53

**Ojeda H, Deloire A, Carbonneau A, Ageorges A, Romieu C.** Berry development of grapevines: Relations between the growth of berries and their DNA content indicate cell multiplication and enlargement. *Vitis* 1999, 38:145-150

**Ollat N, Diakou-Verdin P, Carde J-P, Barrieu F, Gaudillère J-P, Moing A.** Grape berry development: a review. *J Int Sci Vigne Vin* 2002, 36:109-131

**Paran I, van der Knaap E.** Genetic and molecular regulation of fruit and plant domestication traits in tomato and pepper. *J Exp Bot* 2007, 58:3841-52

**Philippe R, Courtois B, McNally KL, Mournet P, El-Malki R, Le Paslier MC, Fabre D, Billot C, Brunel D, Glaszmann JC, This D.** Structure, allelic diversity and selection of *Asr* genes, candidate for drought tolerance, in *Oryza sativa* L. and wild relatives. *Theor Appl Genet* 2010, 121:769-87

**Poupin MJ, Federici F, Medina C, Mattis JT, Timmermann T, Arce-Johnson P.** Isolation of the three grapesub-lineages of B-class MADS-box TM6, PISTILLATA and APETALA3 genes which are differentially expressed during flower and fruit development. *Gene* 2007, 404:10-24

**Pritchard JK, Stephens M, Donnelly P.** Inference of population structure using multilocus genotype data. *Genetics* 2000, 155:945-959

**Ranc.** Analyse du polymorphisme moléculaire de gènes de composantes de la qualité des fruits dans les ressources génétiques sauvages et cultivées de tomate; recherche d'association genes/QTL. Thesis of University of Montpellier, France, 2010, pp273.

**Reid KE, Olsson N, Schlosser J, Peng F, Lund ST.** An optimized grapevine RNA isolation procedure and statistical determination of reference genes for real-time RT-PCR during berry development. *BMC Plant Biology* 2006, 6:27

**Rounsley S, Ditta G, Yanofsky M.** Diverse roles for MADS box genes in Arabidopsis development. *Plant Cell* 1995, 7:1259-1269

**Rozen S, Skaletsky HJ.** Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* 2000, 132:365-86

**Silverstone AL, Tseng TS, Swain SM, Dill A, Jeong SY, Olszewski NE, Sun TP.** Functional analysis of SPINDLY in gibberellin signaling in *Arabidopsis*. *Plant Physiol* 2007, 143:987-1000

**Sreekantan L, Torregrosa L, Fernández L, Thomas MR.** VvMADS9, a class B MADS-box gene involved in grapevine flowering, shows different expression patterns in mutants with abnormal petal and stamen structures. *Funct Plant Biol* 2006, 33: 877-886

**Staden, R.** The Staden Sequence Analysis Package. *Mol Biotech* 1996, 5:233-241

**Suzuki M, Kamide Y, Nagata N, Seki H, Ohyama K, Kato H, Masuda K, Sato S, Kato T, Tabata S, Yoshida S, Muranaka T.** Loss of function of 3-hydroxy-3-methylglutaryl coenzyme A reductase 1 (HMG1) in Arabidopsis leads to dwarfing, early senescence and male sterility, and reduced sterol levels. *Plant J* 2004, 37:750-61

**Takahashi M, Matsuda F, Margetic N, Lathrop M.** Automated identification of single nucleotide polymorphisms from sequencing data. *J Bioinf Comput Biol* 2003, 1:253-265

**Takeuchi T, Watanabe Y, Takano-Shimizu T, Kondo S.** Roles of jumonji and jumonji family genes in chromatin regulation and development. *Dev Dyn* 2006, 235:2449-59

**TanksleySD.** The genetic, developmental and molecular bases of fruit size and shape variation in tomato. *Plant Cell* 2004, 16:181-189

**Tamaru H.** Confining euchromatin/heterochromatin territory: jumonji crosses the line. *Genes Dev* 2010, 24:1465-78

**Terrier N, Glissant D, Grimplet J, Barrieu F, Abbal P, Couture C, Ageorges A, Atanassova R, Léon C, Renaudin JP, Dédaldéchamp F, Romieu C, Delrot S, Hamdi S.** Isogene specific oligo arrays reveal multifaceted changes in gene expression during grape berry (*Vitis vinifera* L.) development. *Planta* 2005, 222:832-847

**Tsukaya H.** Organ shape and size: a lesson from studies of leaf morphogenesis. *Curr Opin Plant Biol* 2003, 6:57-62

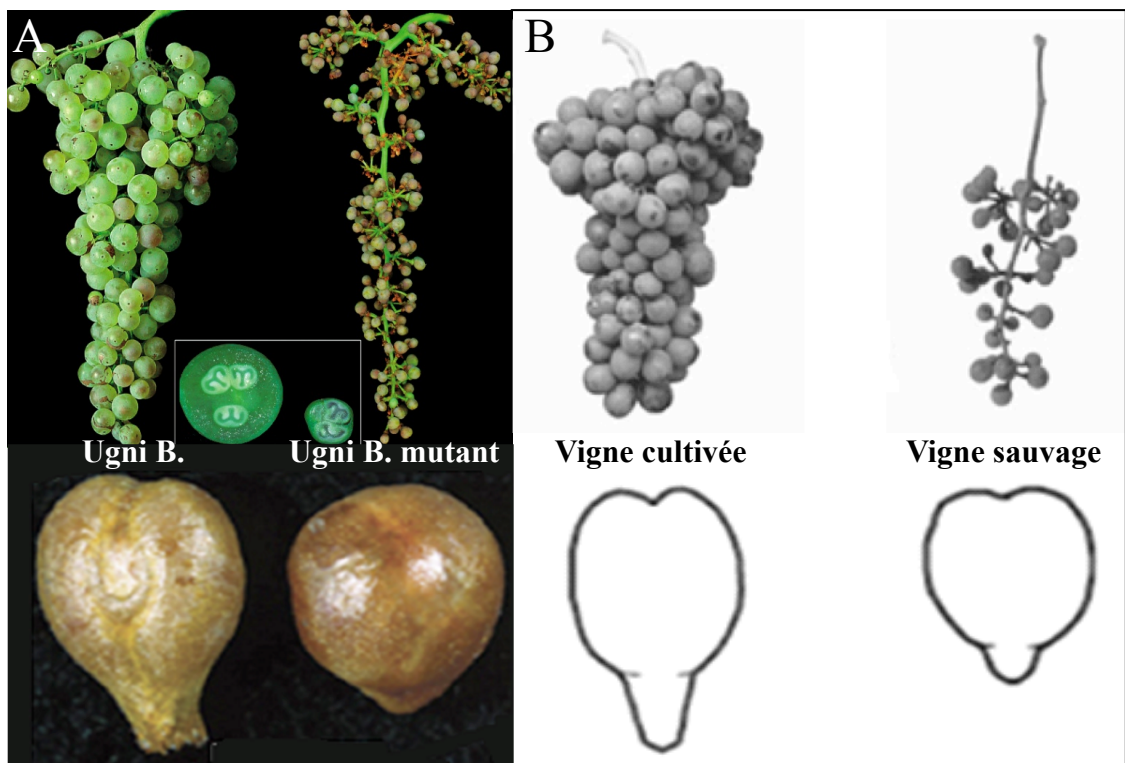
**Yamane K, Tateishi K, Klose RJ, Fang J, Fabrizio LA, Erdjument-Bromage H, Taylor-Papadimitriou J, Tempst P, Zhang Y.** PLU-1 is an H3K4 demethylase involved in transcriptional repression and breast cancer cell proliferation. *Mol Cell* 2007, 25:801-12

**Xiao H, Jiang N, Schaffner E, Stockinger EJ, van der Knaap E.** A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. *Science* 2008, 31

## **5. Recherche de traces de domestication en relation avec la variation de la taille de la baie chez *V. vinifera*: étude du locus *Fleshless berry***



*Photographie d'une grappe de raisins à véraison*



**Figure 5.1:** Différences morphologiques de la baie et du pépin entre (A) l'Ugni Blanc et l'Ugni Blanc mutant et (B) la vigne cultivée et sauvage (d'après Fernandez et al, 2006b et This et al, 2006).

## 5.1. Introduction

Un mutant naturel « *Fleshless berry* » (*Flb*) de la variété Ugni Blanc (Figure 5.1; Fernandez et al, 2006b) a été découvert sur une branche mutée de ce cultivar. Ce mutant est un outil précieux pour l'étude de la variation de la taille du fruit car il présente une altération des divisions des cellules du mésocarpe interne qui sont à l'origine d'environ 90% de volume de la baie chez la vigne cultivée. Par ailleurs, il présente une forme de pépin circulaire caractéristique des vignes sauvages alors que ceux des vignes cultivées sont allongés (Figure 5.1; This et al, 2006). De par leur taille réduite et leurs pépins ronds, les baies du mutant *Flb* ressemblent beaucoup à celles de la vigne sauvage, ce qui nous a conduit à émettre l'hypothèse que des mutations dans le gène *Flb* pourraient avoir été sélectionnées pendant la domestication de la vigne et avoir permis le passage de petites baies avec peu de pulpe à des baies plus charnues. Le processus de sélection et en particulier la domestication peut aboutir à une modification de la diversité génétique des plantes cultivées par rapport à leurs progéniteurs sauvages (Wang et al, 1999; Yamasaki et al, 2007). On s'attendrait donc, sous l'hypothèse précédente, à observer une différence de diversité nucléotidique dans la région du gène *Flb* entre un panel de génotypes sauvages et cultivés.

La mutation *Flb* ségrège comme un gène unique, et l'allèle mutant est dominant. Il a été cartographié à l'aide d'une population de pleins frères de 71 individus, dérivée du croisement Chardonnay x Ugni Blanc mutant dans une région de 5,6 cM sur l'extrémité supérieure du groupe de liaison 18 (Fernandez et al, 2006c). Dans le cadre d'une collaboration entre l'UMR AGAP (INRA Montpellier; A. N'Diaye, L. Torregrosa, A. Bouquet) et le CSIRO (J. Chaïb, MR. Thomas), une nouvelle population à cycle court de développement (1 an de graines à graines; Boss et Thomas, 2002) Picovine x Ugni Blanc mutant de 478 individus a été développée et phénotypée pour la cartographie fine du locus *Flb* (Chaïb et al, 2010). L'objectif d'une partie de ma thèse présentée dans ce chapitre a été (1) de contribuer à la cartographie fine du gène en développant des marqueurs SNP chez l'Ugni Blanc mutant dans la région du gène, et (2) d'analyser la diversité nucléotidique de cette région chez des génotypes sauvages et cultivés afin d'y rechercher d'éventuelles traces de sélection, permettant de tester l'hypothèse du rôle de *Flb* lors de la domestication. De plus, cette dernière approche pourrait permettre de restreindre l'intervalle physique dans lequel se trouve le gène *Flb*.

L'ensemble de ce travail a également été réalisé dans le cadre d'un projet innovant,

SNPGrpMap, coordonné par R. Bacilieri (UMR AGAP, INRA de Montpellier) qui visait à re-séquencer par la technique Sanger 1100 fragments répartis sur le génome de la vigne dans une collection de 28 *V. vinifera* cultivées, 7 *V. vinifera* sauvages et 12 représentants d'autres espèces du genre *Vitis*. Dans ce projet, les amorces PCR, leur test génomique et la première amplification ont été réalisés à l'URGV, puis le séquençage a été réalisé dans l'unité Etude du Polymorphisme des Génomes Végétaux (EPGV, INRA d'Evry; R. Bounon, M-C. Le Paslier, D. Brunel). L'analyse des séquences a ensuite été partagée entre les équipes. De même, le génotypage des SNP a été réalisé sur la plateforme de l'EPGV. Enfin, j'ai également été amenée à collaborer avec la plateforme de cytologie moléculaire de l'UMR APBV (INRA de Rennes; M-S. Vernerey, O. Coriton) pour des expériences de BAC-FISH.

Les résultats que j'ai obtenu sont présentés dans la première partie de ce chapitre sous la forme d'un article publié dans *BMC Plant Biology*. Par la suite, dans un souci de validation des résultats obtenus, un ensemble de SNP bien répartis le long de la région *Flb* a été génotypé sur une collection beaucoup plus importante de vignes sauvages et cultivées. Les résultats complémentaires ainsi obtenus sont présentés dans la seconde partie de ce chapitre, suivie d'une conclusion de celui-ci.

## **5.2. Analyse de la diversité génétique au locus *Flb***

L'analyse de la diversité génétique du locus *Fleshless berry* dans un panel de vignes cultivées et sauvages ainsi que chez l'Ugni Blanc mutant, l'Ugni Blanc sauvage et le Chardonnay, est présentée sous la forme d'un article qui a été accepté le 22 décembre 2010 dans *BMC Plant Biology*. Les figures et tableaux supplémentaires de l'article sont regroupés dans l'annexe 3.



## Patterns of sequence polymorphism in the *Fleshless berry* locus in cultivated and wild *Vitis vinifera* L. accessions

Cléa Houel<sup>1</sup>, Rémi Bounon<sup>1,2</sup>, Jamila Chaïb<sup>3</sup>, Cécile Guichard<sup>1</sup>, Jean-Pierre Péros<sup>4</sup>, Roberto Bacilieri<sup>4</sup>, Alexis Dereeper<sup>4</sup>, Aurélie Canaguier<sup>1</sup>, Thierry Lacombe<sup>4</sup>, Amidou N'Diaye<sup>4</sup>, Marie-Christine Le Paslier<sup>2</sup>, Marie-Stéphanie Vernerey<sup>5,6</sup>, Olivier Coriton<sup>5</sup>, Dominique Brunel<sup>2</sup>, Patrice This<sup>4</sup>, Laurent Torregrosa<sup>4</sup>, Anne-Françoise Adam-Blondon<sup>1</sup>

<sup>1</sup> Unité mixte de Recherche en Génomique Végétale (URGV), INRA UEVE ERL CNRS, 2 rue Gaston Crémieux, 91 057 Evry cedex, France

<sup>2</sup> Unité INRA Etude du Polymorphisme des Génomes Végétaux (EPGV), 2 rue Gaston Crémieux, 91 057 Evry cedex, France

<sup>3</sup> CSIRO Plant Industry, PO BOX 350, Glen Osmond SA 5064, Australia

<sup>4</sup> Unité mixte de Recherche Diversité et Adaptation des Plantes Cultivées (DiaPC), INRA SupAgro, 2 place Pierre Viala, 34 060 Montpellier Cedex, France

<sup>5</sup> Unité mixte de Recherche Amélioration des Plantes et Biotechnologies Végétales (APBV), INRA Agrocampus Rennes, Plate-forme cytologique moléculaire, 35 653 Le Rheu Cedex, France

<sup>6</sup> Unité mixte de Recherche Biologie et Génétique des Interactions Plantes-Agents Pathogènes (BGPI), INRA SupAgro CIRAD, 2 place Pierre Viala, 34 060 Montpellier Cedex, France

### Abstract

**Background:** Unlike in tomato, little is known about the genetic and molecular control of fleshy fruit development of perennial fruit trees like grapevine (*Vitis vinifera* L.). Here we present the study of the sequence polymorphism in a 1 Mb grapevine genome region at the top of chromosome 18 carrying the *Fleshless berry* mutation (*Flb*) in order, first to identify SNP markers closely linked to the gene and second to search for possible signatures of domestication.

**Results:** In total, 62 regions (17 SSR, 3 SNP, 1 CAPS and 41 re-sequenced gene fragments) were scanned for polymorphism along a 3.4 Mb interval (85,127-3,506,060 bp) at the top of the chromosome 18, in both *V. vinifera* cv. Chardonnay and a genotype carrying the *Flb* mutation, *V. vinifera* cv. Ugni Blanc mutant. A nearly complete homozygosity in Ugni Blanc (wild and mutant forms) and an expected high level of heterozygosity in Chardonnay were revealed. Experiments using qPCR and BAC FISH confirmed the observed homozygosity. Under the assumption that *Flb* could be one of the genes involved into the domestication syndrome of grapevine, we sequenced 69 gene fragments, spread over the *Flb* region, representing 48,874 bp in a highly diverse set of cultivated and wild *V. vinifera* genotypes, to

identify possible signatures of domestication in the cultivated *V. Vinifera* compartment. We identified eight gene fragments presenting a significant deviation from neutrality of the Tajima's D parameter in the cultivated pool. One of these also showed higher nucleotide diversity in the wild compartments than in the cultivated compartments. In addition, SNP significantly associated to berry weight variation were identified in the *Flb* region.

**Conclusions:** We observed the occurrence of a large homozygous region in a non-repetitive region of the grapevine otherwise highly-heterozygous genome and propose a hypothesis for its formation. We demonstrated the feasibility to apply BAC FISH on the very small grapevine chromosomes and provided a specific probe for the identification of chromosome 18 on a cytogenetic map. We evidenced genes showing putative signatures of selection and SNP significantly associated with berry weight variation in the *Flb* region. In addition, we provided to the community 554 SNP at the top of chromosome 18 for the development of a genotyping chip for future fine mapping of the *Flb* gene in a F2 population when available.

## **Background**

Berry size is an important trait in relation to both yield (table grapes) and quality (wine grapes). Indeed, the flavor in wine results from the ratio of skin to flesh, the former being the source of most aromatic and tannins compounds, the second providing the organic acids and the sugars [1].

The genetic and molecular basis of fleshy fruit size variation have been studied in depth in tomato during the last two decades, using a large panel of diverse resources that made tomato a model species for fleshy fruit crops [2,3]. Introgression lines between wild and cultivated genotypes [4,5,6], near isogenic lines (NILs) [7] and artificial or natural mutants [2,8] have been created and used to study the genetic basis of fruit size variation showing that a large part of it is controlled by less than ten loci. The physiological mechanisms involved have been related to the control of (i) the cell number in the pericarp, as for the *fw2.2* locus [9,2,10], (ii) the locule number [2,11], (iii) the late endo-reduplication in pericarp cells [12,10] and (iv) the cell wall plasticity in relation to the cell expansion [10]. All these advances in tomato are useful to assist the study of similar trait in other crops with fleshy fruits, less amenable to genetic studies, such as perennial fruit trees. Indeed, encouraging results have already shown synteny within Solanaceae species for Quantitative Trait Loci (QTL) controlling fruit size

[2]. However, the degree of transferability of knowledge from tomato to non-Solanaceae species remains an open question.

Like tomato, grapevine (*Vitis vinifera* L.) produces fleshy fruits and a large difference in fruit size between wild and cultivated genotypes can be observed [13]. Indeed the wild *V. vinifera* genotypes produce mature berries weighting less than 1g while berries of some table grape varieties can weigh 10g and more [14]. The growth of a grapevine berry roughly follows the same pattern as for tomato fruit: the first phase of fruit development is due to both cell multiplication and cell expansion, followed by a lag phase corresponding to a major cell metabolic shift and a second phase of fruit growth, mostly explained by cell expansion but without evidence of endoreduplication [15]. The genetic analysis of grape berry size variation is more difficult than in tomato, due to the long biological cycle of the plant, to the high level of heterozygosity of the genome and to the large field area usually required for plant growth, which makes experiments in controlled environment more costly [16,17,18,19]. In addition, berry size studies have often been performed on population segregating for seedlessness, with a strong negative correlation between the two traits: the seedless berries are in average smaller than the seeded berries [16,17,18,19]. Up to now, it has not been possible to establish the relationship between QTL for berry size and processes like cell multiplication or cell enlargement.

A natural mutant of *V. vinifera* cv. Ugni Blanc, which produces fleshless berries similar to those observed in wild genotypes, was identified as an opportunity to get insights into the control of berry development and berry size [20]. It has been shown that the drastic phenotypic changes observed in berry development are controlled by a dominant mutation in the *Fleshless berry* (*Flb*) gene [21]. Like the *fw2-2* gene in tomato, the *Flb* gene impairs cell divisions in the developing ovaries [21]. The closest genetic marker linked to the *Flb* mutation defines a 6 cM region located at the top the chromosome 18 that corresponds to a physical distance of 948 kb according to the last version of the grapevine genome assembly (<http://urgi.versailles.inra.fr/projects/Vitis/ressources.php>); in this region, no homologue to the *fw2-2* has been identified. Considering the importance of berry size for wine quality, a fine mapping of the *Flb* mutation was thus started for its molecular identification.

Here we describe our efforts in reducing the genome interval of the region carrying the *Flb* mutation. We first started by a classical genetic mapping approach. We showed that the mutation is located on a completely homozygous portion of chromosome 18 in Ugni Blanc

mutant. No marker could thus be found in coupling with the mutation and the classical approach was abandoned.

We therefore started another approach similar to the one previously proposed for *fw2-2* gene in tomato [9]. Since the berries of Ugni Blanc mutant mimic wild *V. vinifera* berries (both types of berries have little to no flesh and carry round shaped seeds typical of wild genotypes) [13,20,22], we hypothesized that *Flb* gene could have been one of the genes selected during the domestication process of grapevine. If so, a signature of selection or selective sweep could be found around this gene. Under this assumption, we performed a preliminary scan of the sequence polymorphism of the *Flb* region in a collection cultivated and wild grapevine genotypes.

## **Methods**

### **Plant material**

The genotypes used in the present study were collected in the French National Grapevine Germplasm Collection (Domain of Vassal, Montpellier, France; <http://www1.montpellier.inra.fr/vassal/>) and are listed in additional file 1. Twenty-six of them were chosen to maximize the genetic diversity of the cultivated *V. vinifera* compartment [23]. Seven other genotypes belonging to the wild *V. vinifera* compartment were chosen because they had well characterized wild-type phenotypes as well as wild-type diverse SSR profiles and because they originated from different countries (8500Mtp31 from Tunisia, 8500Mtp9 and 8500Mtp38 from Germany and the rest from France; [additional file 1]). Five genotypes were added to the sample: the inbred line INRA Colmar lignée PN40024 (PN40024; reference genome; maintained at INRA Colmar, France), Chardonnay, Ugni Blanc, Ugni Blanc mutant and Pinot Noir clone ENTAV-INRA777 (PN777; maintained at the French Institute for Grapevine and Wine; Domaine de l’Espiguette, Le Grau du Roi, France). The average berry weight at maturity was measured from 30 berries cut at the pedicel base 40 days after *véraison* [additional file 1].

### **DNA extraction**

Total genomic DNA was extracted from 1 g of young leaves according to the DNeasy Plant Maxi Kit (Qiagen) with the following modifications: 1% polyvinylpyrrolidone (PVP 40 000) and 1% (v/v)  $\beta$ mercaptoethanol were added to buffer AP1. The clarified lysate recovered after filtration with the QIA-shredder Maxi spin column (step 12) was extracted with one volume

of phenol:chloroform:isoamyl alcohol (25:24:1) and then with one volume of chloroform:isoamyl alcohol (24:1). From this step forward, the supernatant was treated following the Qiagen instructions.

### **Gene fragments amplification and sequencing**

Based on the genome annotation provided by Jaillon et al [24], 86 primer pairs were designed using the Primer 3 software v.0.4.0 [25] in order to amplify every 12 kb in the *Flb* region, a gene fragment of approximately 1300 bp [additional file 2]. In order to estimate the nucleotide diversity at the whole genome scale, seventy-seven other primer pairs were designed on genes chosen randomly along the genome, taking care that each chromosome was represented by three to five fragments [additional file 3]. The amplicon sequences were then aligned on the last 12X version of the genome sequence ([http://urgi.versailles.inra.fr/cgi-bin/gbrowse/vitis\\_12x\\_pub/](http://urgi.versailles.inra.fr/cgi-bin/gbrowse/vitis_12x_pub/)) and some of them did not correspond to a gene model anymore. Settings for Primer 3 were: optimum Tm=55°C, minimum Tm=53°C, maximum Tm=57°C, max 5' self complementarity=4, max 3' self complementarity=1. In order to amplify all the genotypes while at the same time detecting a maximum of polymorphism, all the primers were designed in exons at both sides of introns. Universal primers T7/SP6 extensions were added to the primers to allow sequencing. All PCR amplifications were carried out as described by Philippe et al [26].

### **Microsatellite, CAPS and SNP genotyping**

The markers genotyped are listed and described in additional file 2. Cleaved Amplified Polymorphic Sequence (CAPS) genotyping was performed as described by Salmaso et al [27]. The Australian Genome Research Facility (AGRF) carried out Simple Sequence Repeats (SSR) and Single Nucleotide Polymorphism (SNP) analysis. SNP were scored using the MassARRAY® iPLEXGold assay with MALDI-TOF MS detection (Sequenom) and SSR analysis was performed as previously described by Thomas et al [28].

### **Quantitative PCR assay**

Two primer pairs were designed to amplify genomic DNA. The first pair (TCTGATGCGATGTTAGTGGT and TCTGGTATTGGCGTTGG) targeted a unique gene (*FL*) in the *Flb* region (gene ID GSVIVG01013466001). The second pair (AACTGGATTGAAGGGCGTGG and AGGTTCTTGAGCATGTTAAGC) targeted the 3-



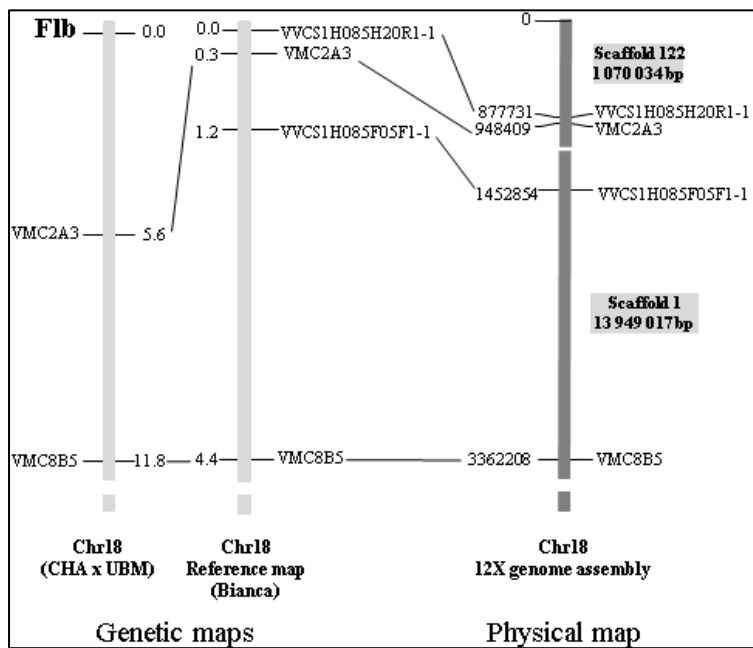
*hydroxy-3-methylglutaryl-coenzyme A reductase (HMGCoA)* gene family, which members are respectively located on the chromosomes 4, 3 and 18 (gene id GSVIVG01026444001, GSVIVG011023852001, GSVIVG01013435001). Real-time PCR conditions were conducted as described by Reid et al [29], with half quantity of PCR mix and of DNA. The PCR efficiencies were determined for each gene and were 92.3% and 97.2% for *FL* and *HMGCoA* respectively. In order to compare the initial DNA quantity between genotypes in the *Flb* region, the DNA quantity based on *FL* gene data was normalised using the DNA quantity of *HMGCoA* genes as a reference.

### **BAC-FISH assay**

Roots tips of 0.5-1.5 cm length were treated in the dark with 0.04% 8-hydroxyquinoline for 2 h at 4°C followed by 2 h at room temperature to accumulate metaphases. They were then fixed in 3:1 ethanol-glacial:acetic acid for 12 hours at 4°C and stored in ethanol 70 % at -20 °C. They were washed in 0.01 M citric acid-sodium citrate pH 4.5 buffer for 15 min and then digested in a solution of 5% Onozuka R-10 cellulase (Sigma), 1% Y23 pectolyase (Sigma) at 37 °C for 1 h. Digested root tips were then carefully washed with distilled water for 2 h. One root tip was transferred to a slide and macerated in a drop of 3:1 fixation solution (ethanol-glacial:acetic acid). Chromosome spreads were prepared for hybridization as described by Leflon et al [30]. VV40024H140P14 Bacterial Artificial Chromosome (BAC) clone (available at <http://cnrgv.toulouse.inra.fr>) was labelled by random priming with biotin-14-dUTP (Invitrogen). The ribosomal probe used, as a control of hybridization, was pTa-71 which contains a 9 kb *EcoRI* fragment of ribosomal DNA repeat unit (rDNA 18S-5.8S-26S genes and spacers) isolated from *Triticum aestivum* [31]. The probe pTa-71 was labelled with Alexa-488 dUTP (Invitrogen) by random priming. Fluorescence *In Situ* Hybridization (FISH) experiments and capture of fluorescence images were done as described by Leflon et al [30].

### **Sequence data analysis, estimation of parameters of diversity and linkage disequilibrium**

Raw data were aligned and trimmed using either the Genalys v.2.8.3b software for Macintosh [32] or the Staden software v.2.0.0 [33]. They were manually edited and INsertions/DEletions (INDEL) were added when needed. Single Nucleotide polymorphisms (SNP) were detected, confirmed, and imported into the SNIPlay database (<http://sniply.cirad.fr>). Nucleotide diversity ( $\pi$ ), number of segregating sites ( $\theta$ ), number of haplotype (H), haplotype diversity (Hd), and Tajima's D test of neutral evolution [34] were



**Table 1.** Marker polymorphism observed between cultivars Chardonnay and Ugni Blanc mutant on the top the chromosome 18 (12X genome assembly)

Position on the chromosome 18 (bp)			Marker name	Marker type	Chardonnay <sup>§</sup>	Ugni Blanc mutant <sup>§</sup>
Scaffold	Start	End				
122	212555	212699	VS50	SSR	H	h
122	213864	214083	VS51	SSR	H	h
122	226346	226575	VS52	SSR	h	h
122	230668	230769	VS53	SSR	H	h
122	308176	308256	1036L11F	SNP	h	h
122	321067	321135	VMC3E5	SSR	h	h
122	388123	388423	VVIN03	SSR	h	h
122	423185	423271	1038A12F	SNP	h	h
122	494374	494464	VS54	SSR	H	h
122	497723	498045	IN0954	CAPS	h	h
122	670015	670200	VS56	SSR	h	h
122	804498	804634	VS57	SSR	H	h
122	877751	878077	WCS1H085H20R1-1	SSR	h	h
122	895761	895846	1073P15R	SNP	h	h
122	901775	901934	VS58	SSR	H	h
122	948267	948387	VMC2A3*	SSR	H	h
1	1226489	1226647	WCS1H066N21R1-1	SSR	H	h
1	1297892	1298020	C011	SSR	H	h
1	1452854	1453153	WCS1H085F05F1-1	SSR	H	H
1	2912753	2913088	WVIB31	SSR	h	h
1	3505999	3506060	WVIV16	SSR	H	h

<sup>§</sup> H: for heterozygous marker and h: for homozygous marker

\* From Fernandez et al [21]



obtained for each gene fragment using the DnaSp V5.10 software (<http://www.ub.edu/dnasp/>). Eventually, the total value of each parameter was calculated as a weighted average for the whole data set. As all the gene fragments along the *Flb* region were separated in average by 12 kb (from 3 to 57 kb), it was not possible to reconstitute the haplotypes for the entire *Flb* region in order to estimate the Linkage Disequilibrium (LD). Roger and Huff [35] showed that the genotypic correlation coefficient (based on genotypic data) is a good estimator of the haplotypic correlation coefficient. LD was therefore estimated over the entire studied region as the square of the genotypic Pearson correlation coefficient ( $r^2$ ) together with its p-value using a homemade R program. The results were visualised using in homemade Perl scripts.

### **Association genetics**

A structured association test was carried out using TASSEL software ([http://www.maizogenetics.net/index.php?option=com\\_content&task=view&id=89&Itemid=119](http://www.maizogenetics.net/index.php?option=com_content&task=view&id=89&Itemid=119)). The population structure was calculated using STRUCTURE software [36] using the genotypes at 20 SSR markers well spread along the 19 chromosomes (Le Cunff et al, 2008; R. Bacilieri unpublished results; [additional file 1]). A General Linear Model test, which takes into account the structure of the sample, was performed between the SNP markers in the *Flb* region with a allelic frequency >0.05 and the average berry weight at maturity log-transformed. A Bonferroni correction was applied to control false-positives: a SNP marker was declared significant if its Bonferroni P-value was less than 0.05.

## **Results**

### **A 1Mb region at the top of chromosome 18 is homozygous in Ugni Blanc and the *Fleshless berry* mutant**

The *Flb* mutation was localised by Fernandez et al [21] at the top of chromosome 18, above the marker VMC2A3 and VMC8B5 on the consensus map of a progeny of Chardonnay by Ugni Blanc mutant. However, the *Flb* locus was mapped indirectly relative to VMC2A3 that segregated in Chardonnay and not in Ugni Blanc mutant. For the purpose of finding polymorphic markers in Ugni Blanc mutant above VMC2A3, we aligned the genetic map to the grapevine reference genome sequence [24] in order to identify SSR and SNP markers segregating in the Ugni Blanc mutant. This region corresponded to 948 kb on chromosome 18 (upper part of scaffold 122; Figure 1) where 100 predicted genes were proposed by the automatic annotation.

**Table 2.** Sequence polymorphism observed between cultivars Chardonnay and Ugni Blanc mutant on the top of the chromosome 18 (12X genome assembly)

Position on the 12x genome assembly (bp)			Fragment name	Number of extremities sequenced	Sequence Length		Homozygous polymorphic sites between Chardonnay and Ugni Blanc mutant		Chardonnay: number of heterozygous		Ugni Blanc mutant: number of heterozygous	
Scaffold	Start	End			Chardonnay	Ugni Blanc mutant*	SNP	INDEL	SNP	INDEL	SNP	INDEL
122	85127	85871	WC2982A	2	1,553	1,553	0	0	2	0	0	0
122	161551	161929	WV05806A	2	1,168	1,168	0	0	0	0	0	0
122	211001	211674	WC2974A	1	969	969	0	0	4	0	0	0
122	261445	262084	WV05805A	2	1,562	1,562	0	0	10	0	0	0
122	299201	299664	WC2967B	2	911	911	0	0	2	0	0	0
122	321452	321822	WV05803A	2	1,077	1,077	0	0	7	0	0	0
122	372496	372799	WV05800A	2	683	683	1	0	2	0	0	0
122	382744	382940	WC2956A	2	452	876	0	0	2	1	0	0
122	399382	399793	WC2953A	2	769	1,505	0	0	0	1	0	0
122	429725	431077	WV05799A	2	1,539	1,539	2	0	1	0	0	0
122	497378	497760	WC2942A	2	933	1,255	0	0	5	1	0	0
122	510613	510723	WV05798A	1	1,464	1,464	0	0	4	0	0	0
122	549494	550104	WV05796A	2	909	909	1	0	2	0	0	0
122	615081	615296	WV05793A	1	407	1,057	0	0	1	1	0	0
122	668381	668534	WV05788A	1	914	1,053	0	0	2	1	0	0
122	702907	703637	WV05785A	1	1,090	1,446	0	0	0	0	0	0
122	776756	777088	WV05782A	1	413	1,434	2	0	2	1	0	0
122	818661	819292	WV05781A	1	830	1,495	0	0	1	1	0	0
122	898379	898848	WV05779A	1	1,565	1,565	1	0	1	0	0	0
122	928463	929045	WV05777A	2	1,030	1,520	1	0	1	1	0	0
<b>122</b>	<b>949921</b>	<b>950653</b>	<b>WV05775A</b>	<b>1</b>	<b>755</b>	<b>1,116</b>	<b>1</b>	<b>0</b>	<b>1</b>	<b>0</b>	<b>0</b>	<b>0</b>
<b>122</b>	<b>1009539</b>	<b>1010696</b>	<b>VVC2869A</b>	<b>1</b>	<b>137</b>	<b>498</b>	<b>0</b>	<b>1</b>	<b>0</b>	<b>1</b>	<b>0</b>	<b>0</b>
<b>122</b>	<b>1054896</b>	<b>1056021</b>	<b>VVC2865A</b>	<b>1</b>	<b>136</b>	<b>136</b>	<b>0</b>	<b>0</b>	<b>1</b>	<b>1</b>	<b>0</b>	<b>0</b>
<b>1</b>	<b>1084916</b>	<b>1085337</b>	<b>VVC15574A</b>	<b>2</b>	<b>161</b>	<b>161</b>	<b>0</b>	<b>0</b>	<b>3</b>	<b>0</b>	<b>0</b>	<b>0</b>
<b>1</b>	<b>1098027</b>	<b>1099246</b>	<b>VVC15572A</b>	<b>2</b>	<b>1,700</b>	<b>1,700</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
<b>1</b>	<b>1104978</b>	<b>1106307</b>	<b>VVC15571A</b>	<b>2</b>	<b>435</b>	<b>986</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
				41	23,562	29,638	9	1	54	10	0	0

\* The column Ugni Blanc mutant stands for both Ugni Blanc and Ugni Blanc mutant, as no differences were observed between them.  
The table lines in bold characters correspond to the sequence fragments below marker VMC2A3.

First, 17 SSR, three SNP and one CAPS markers were either developed or retrieved from published genetic maps [37,38,39,40,41,42] along scaffold 122 and the beginning of scaffold 1, above and below VMC2A3 [additional file 2]. All primer pairs successfully amplified Chardonnay and Ugni Blanc mutant genomic DNAs. One of them (VVS55), not targeting a single locus, was discarded. Chardonnay was heterozygous for ten of the 20 remaining markers, while Ugni Blanc mutant was always homozygous except for VVCS1H085F05F1-1, which is located after VMC2A3 (Table 1).

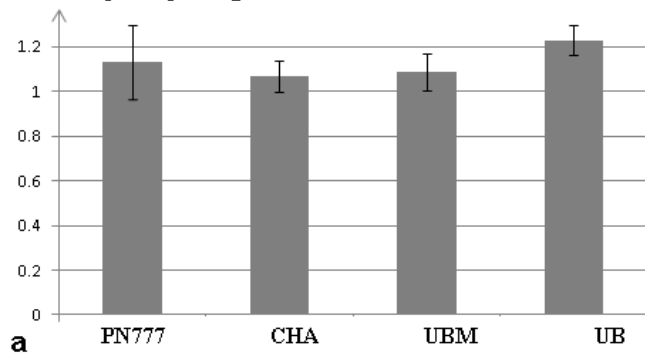
In order to find new heterozygous markers in the *Flb* region, we decided to carry out a re-sequencing approach. Thirty primer pairs were designed along this region [additional file 2], 24 above the SSR marker VMC2A3 and six below. Twenty out of 24 primer pairs (above VMC2A3) successfully amplified the PN40024 genomic DNA and were thus used to sequence the corresponding gene fragments in Chardonnay, Ugni Blanc and Ugni Blanc mutant. We decided to sequence also Ugni Blanc in order to check if the homozygosity of the *Flb* region was specific to the mutant or already present in the wild type.

The 26 fragments of 1300 bp in average were sequenced either only in forward or also in reverse direction, leading to 41 sequences of 161 to 1700 bp long (Table 2), heterozygous INDEL or short repeats leading to the shorter sequences. In total, we analyzed 23,562 bp in Chardonnay and 29,638 bp in Ugni Blanc and Ugni Blanc mutant. This difference was the first observed contrast between Chardonnay and Ugni Blanc, due to a different level of heterozygosity. Comparing the sequences of Chardonnay and Ugni Blanc, 74 polymorphisms were identified (63 SNP and 11 INDEL). Out of these, 10 differences correspond to homozygous SNP or INDEL in both samples, while 64 differences correspond to SNP heterozygous in Chardonnay and homozygous in Ugni Blanc. No heterozygous SNP or INDEL were observed in Ugni Blanc and its mutant; we deduced that the homozygosity of this region derived from Ugni Blanc. Only Ugni Blanc mutant sequences were considered in the subsequent experiments.

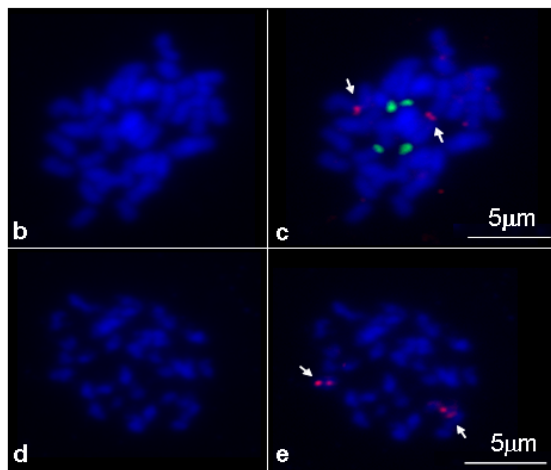
In total, 62 regions (17 SSR, three SNP, one CAPS and 41 re-sequenced gene fragments) were scanned for polymorphism both in Chardonnay and Ugni Blanc mutant along a 3.4 Mb interval (85,127-3,506,060 bp) in the *Flb* region (scaffold 122 and the beginning of scaffold 1). This allowed showing a nearly complete homozygosity in Ugni Blanc mutant and as expected, a high level of heterozygosity in Chardonnay.

To discriminate between a complete homozygosity of Ugni Blanc mutant and a large deletion of the *Flb* region, two experiments were realized. First, a quantitative PCR (qPCR) assay was

Normalized number of *FL* gene copy per haploid genome



**a**

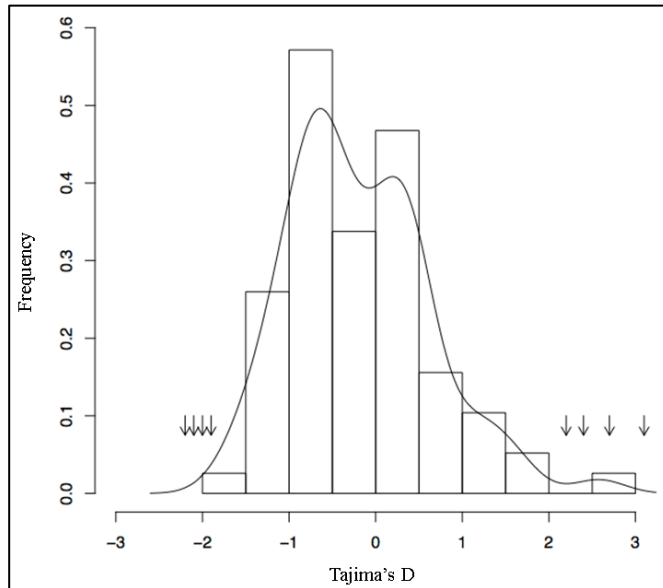


**b**

**c**

**d**

**e**



**Figure 3.** Distribution in cultivated grapevines of the Tajima's D value calculated from the 77 genes randomly distributed across the genome. The arrows correspond to the Tajima's D value from the eight gene fragments in the *Flb* region with a significant deviation of the Tajima's D value from neutrality.

**Figure 2.** Experimental demonstration of homozygosity of the *Flb* region in Ugni Blanc mutant. **(a)** Estimation of the number of *FL* gene copy after normalization in Pinot Noir (PN777), Chardonnay (CHA) Ugni Blanc (UB) and Ugni Blanc mutant (UBM). **(b-c)** Double fluorescence in situ hybridization (FISH) with BAC clone VV40024H140P14 (red) and pTa-71 (green) as a control, on mitotic metaphase chromosomes of Ugni Blanc mutant and **(d-e)** FISH signals of BAC clone VV40024H140P14 (red) on mitotic chromosomes of Pinot Noir (PN777) are indicated with arrows. Chromosomes were counterstained with DAPI (blue).

Table 3. Nucleotide diversity in the wild and cultivated *V. vinifera* genotypes for the gene fragments along the *Flb* region presenting a significant deviation from neutrality of the Tajima's D parameter.

Fragment	Start 12x	Wild				Domesticated				Wine				Table			
		$\pi$	$\pi$ standard error	$\pi$	$\pi$ standard error	Tajima's D <sup>5</sup>	$\pi$	$\pi$ standard error	Tajima's D <sup>5</sup>	$\pi$	$\pi$ standard error	Tajima's D <sup>5</sup>	$\pi$	$\pi$ standard error	Tajima's D <sup>5</sup>		
VVC2981A	94259	0.0007	0.0001	0.0014	0.0005	-2.0*	0.0013	0.0003	-0.5	0.0016	0.0008	-2.2**					
VVC2946A	444180	0.0030	0.0004	0.0043	0.0011	-1.3	0.0035	0.0017	-2.2**	0.0047	0.0016	-1.2					
VV05791A	638081	0	0	0.0040	0.0002	0.6	0.0034	0.0005	1.1	0.0037	0.0003	2.4*					
VVC2897A	682572	0.0173	0.0062	0.0044	0.0017	-2.1*	0.0070	0.0042	-2.1	0.0027	0.0004	-0.7					
VV05785A	702907	0.0008	0.0004	0.0003	0.0002	-1.9*	0.0003	0.0002	-1.5	0.0002	0.0001	-1.5					
VVC2901A	742593	0.0082	0.0035	0.0156	0.0005	3.1**	0.0157	0.0011	2.4*	0.0144	0.0079	2.6**					
VVC2885A	746740	0.0115	0.0037	0.0159	0.0007	2.7**	0.0167	0.0012	2.8**	0.0145	0.0020	1.5					
VVC2892A	808955	0.0009	0.0003	0.0109	0.0007	2.2 *	0.0102	0.0011	2.0	0.0119	0.0009	2.0					

<sup>5</sup> \* 0.01<P-value < 0.05 and \*\* 0.001<P-value < 0.01

<sup>5</sup> \* 0.01<P-value < 0.05 and \*\* 0.001<P-value < 0.01

performed on genomic DNA from Ugni Blanc, Ugni Blanc mutant, Chardonnay, PN777 and PN40024 as controls. No difference in the estimation of the initial DNA quantity was observed when amplifying with primer pair *FL*, which targeted a gene in the *Flb* region and the other primer pair *HMGC $\alpha$ A*, which targeted three loci elsewhere in the genome (Figure 2a; [additional file 4]). This indicated that this region is homozygous and not deleted in Ugni Blanc or Ugni Blanc mutant. The second experiment consisted in a FISH experiment with a BAC clone (VV40024H140P14) localized specifically in the *Flb* region using mitotic metaphase chromosomes of Ugni Blanc mutant and PN777 as control. Chromosomes were counter stained with DAPI (Figure 2b-e) and FISH signals corresponding to VV40024H140P14 were detected on two homologous chromosomes in both PN777 and Ugni Blanc mutant (Figure 2c and 2e respectively), which confirmed that the *Flb* region was not deleted in Ugni Blanc mutant.

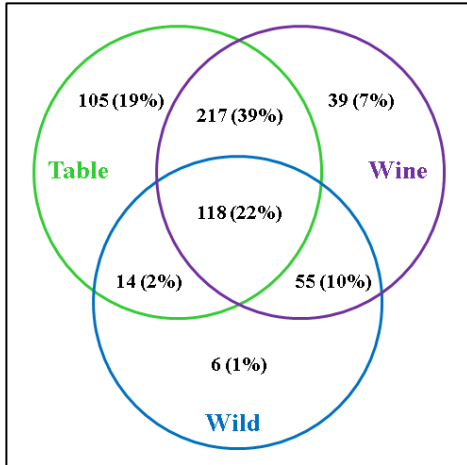
### ***Flb* region showed possible signatures of selection in the cultivated *V. vinifera* compartment**

A fragment every ten to 20 kb, in the 948 kb region above marker VMC2A3 was re-sequenced in a highly diverse set of cultivated *V. vinifera* genotypes [additional file 1], in order to evidence possible traces of selection in the cultivated pool of grapevines.

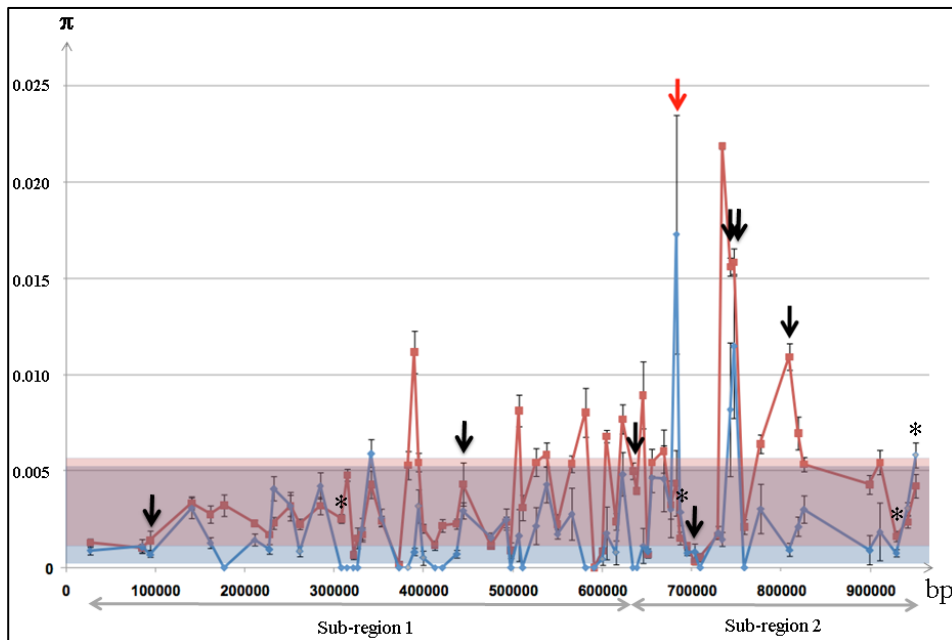
Sixty-three additional primer pairs were developed; two of them being discarded because they did not amplify in PN40024 [additional file 2]. Eighty-two primer pairs (20 targeting fragments before VMC2A3, one targeting a fragment after VMC2A3 described in the former paragraph and 61 newly developed) were thus used to sequence the corresponding gene fragments in 26 cultivated *V. vinifera* and the PN40024 as control [additional file 1]. Each fragment was compared to the 12X version of the genome reference sequence, which allowed us to discard the results obtained for eight and three fragments that appeared to be either part of a false duplication in the 8X version of the genome sequence, or to the same gene in the 12X gene annotation, respectively [additional file 2]. The remaining data, from 69 sequenced regions, consisted in a total of 34,355 kb, 61% (21,161 kb) being located in predicted introns or UnTranslated Region (UTR) and 39% (13,194 kb) in exons [additional file 5]. In parallel, 77 random gene fragments spread all over the genome were chosen in order to estimate the nucleotide diversity over the whole genome, and as control for the effect of selection. These gene fragments represented 48,874 kb of total sequence, 55% (27,018 kb) located in predicted introns or UTR and 45% (21,856 kb) in predicted exons [additional file 3].

**Table 4.** Summary of the sequence polymorphism observed in cultivated and wild *V. vinifera* genotypes for 69 sequence fragments along 948 kb in the *Flb* region and for 77 sequence fragments spread along the whole genome.

	Average number of genotypes/fragment	Average number of segregating sites/fragment	Average number of haplotypes/fragment	Average and standard deviation of $\pi$
<b>Wild (n = 7)</b>				
<i>Flb</i> region	5.9	2.8	2.2	0.0020 +/- 0.0006
Whole genome	6.7	4.8	3.6	0.0027 +/- 0.0025
<b>Cultivated (n = 26)</b>				
<i>Flb</i> region/Wine (n = 15)	10.5	6.2	4.3	0.0035 +/- 0.0007
<i>Flb</i> region/Table (n = 11)	12.5	6.6	4.7	0.0035 +/- 0.0007
<i>Flb</i> region/Wine + Table	24.8	8	5.8	0.0041 +/- 0.0004
Whole genome/Wine + Table	27.0	10.1	8.3	0.0035 +/- 0.0023



**Figure 4.** SNP from the *Flb* region in wild and cultivated grapevines. Venn diagram showing the distribution of the 554 non-redundant SNP found in the 948 kb region at the top of chromosome 18 in the sets of wild and domesticated table and wine *V. vinifera* genotypes.

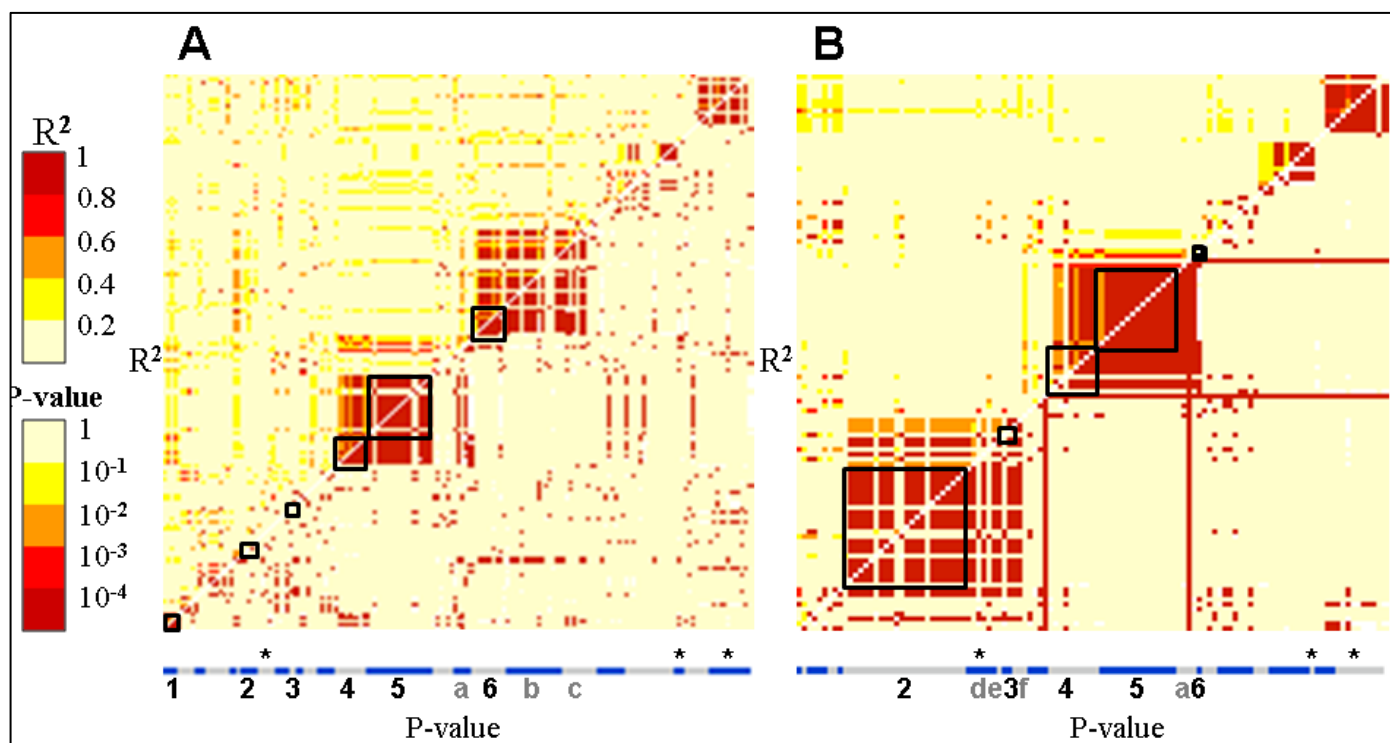


**Figure 5.** Nucleotide diversity in wild and cultivated grapes along the *Flb* region. Nucleotide diversity ( $\pi$ ) in wild (blue line) and cultivated grapes (red line) along the *Flb* region. The standard deviation of the  $\pi$  parameter in the whole genome is represented by a blue and red box for wild and cultivated genotypes respectively. Genes under selection in the cultivated pool of genotypes are indicated with black arrows and the gene under purifying selection showing higher diversity in wild genotypes than in cultivated genotypes with red arrows. The two sub-regions with regard to LD patterns are underlined with grey arrows. Gene fragments with SNP significantly associated with berry weight variation are highlighted with a star.

The Tajima's D parameter, was calculated for the 77 random genes and for the 69 genes from the *Flb* region [additional file 3 and 5]. Eight of 69 sequenced fragments in the *Flb* region showed putative traces of selection evidenced by a Tajima's D parameter significantly deviating from neutrality (Table 3). Moreover, for these fragments, the value of Tajima's D parameter was quite divergent from the average calculated for the 77 random genes ( $-0.1853 \pm 0.8117$ ; [additional file 3]) and were found in the tails of the distribution of Tajima's D value across the genome (for  $\alpha=0.05$ ; Figure 3). A significant negative Tajima's D value, possibly indicative of a purifying selection was observed for four out of the eight gene fragments whereas a significant positive Tajima's D value, possibly indicative of a diversifying selection, was found for the other four (Table 3).

#### **Analysis of the nucleotide diversity along the *Flb* region in a set of cultivated and wild *V. vinifera* genotypes**

The 69 gene fragments from the *Flb* region and the 77 random gene fragments spread all over the genome were sequenced in seven diverse wild *V. vinifera* genotypes, in order to compare the nucleotide diversity in the cultivated and wild pools of genotypes. The diversity parameters calculated for each fragment in the two different subsets of individuals, are presented in additional files 3 and 5 and summarized in Table 4. All the indicators of genetic diversity (number of segregating sites, number of haplotypes, and nucleotide diversity:  $\pi$ ) were higher in average (roughly doubled  $\pi=0.0041$  vs  $0.0020$ ; [additional file 5]) in the whole sample of domesticated genotypes in comparison to the sample of wild genotypes in the *Flb* region. This hold true when each of the wine and table grape sub-compartments of cultivated grapes were compared with the wild compartment, with less unbalanced numbers of individuals in each pairwise comparison (Table 3). Compared to a similar number of re-sequenced fragments spread all over the genome, there was a slightly lower diversity among the wild genotypes in the *Flb* region than in the rest of the genome, which was not the case in the cultivated compartment (Table 4). Moreover, we observed very few specific segregating sites between the wild and the cultivated compartment in the *Flb* region (out of 554 SNP sites, only six were specific to the wild compartment; Figure 4; [additional file 5]). Nucleotide diversity varied along the *Flb* region, also depending on the pool of genotypes considered (Figure 5; [additional file 5; additional file 6]) and was locally higher in the cultivated compartment than in the wild compartment (Figure 5). This probably reflected the fact that 18 out of 69 fragments showed no sequence polymorphism among the wild genotypes [additional



**Figure 6.** Linkage disequilibrium along the second *Flb* sub-region in the cultivated and wild compartments. LD plots on  $R^2$  values (above the diagonal) and associated P-value (below the diagonal) along the second sub-region containing the 6 gene fragments under selection in the cultivated (A) and wild (B) compartments. The gene fragments re-sequenced are represented by alternate grey and blue boxes, which size is proportional to the number of polymorphic SNP used in the LD estimation. The gene fragments under selection are in black boxes and numbered as follows: “1” for VV05791A, “2” for VVC2897A, “3” for VV05785A, “4” for VVC2901A, “5” for VVC2885A and “6” for VVC2892A. The gene fragments in LD with these genes are pointed with small letter “a” for VV05782A, “b” for VV05781A, “c” for VV05780A, “d” for VV05786A, “e” for VVC2907A and “f” for VVC2891A. Gene fragments with SNP significantly associated with berry weight variation are highlighted with a star.

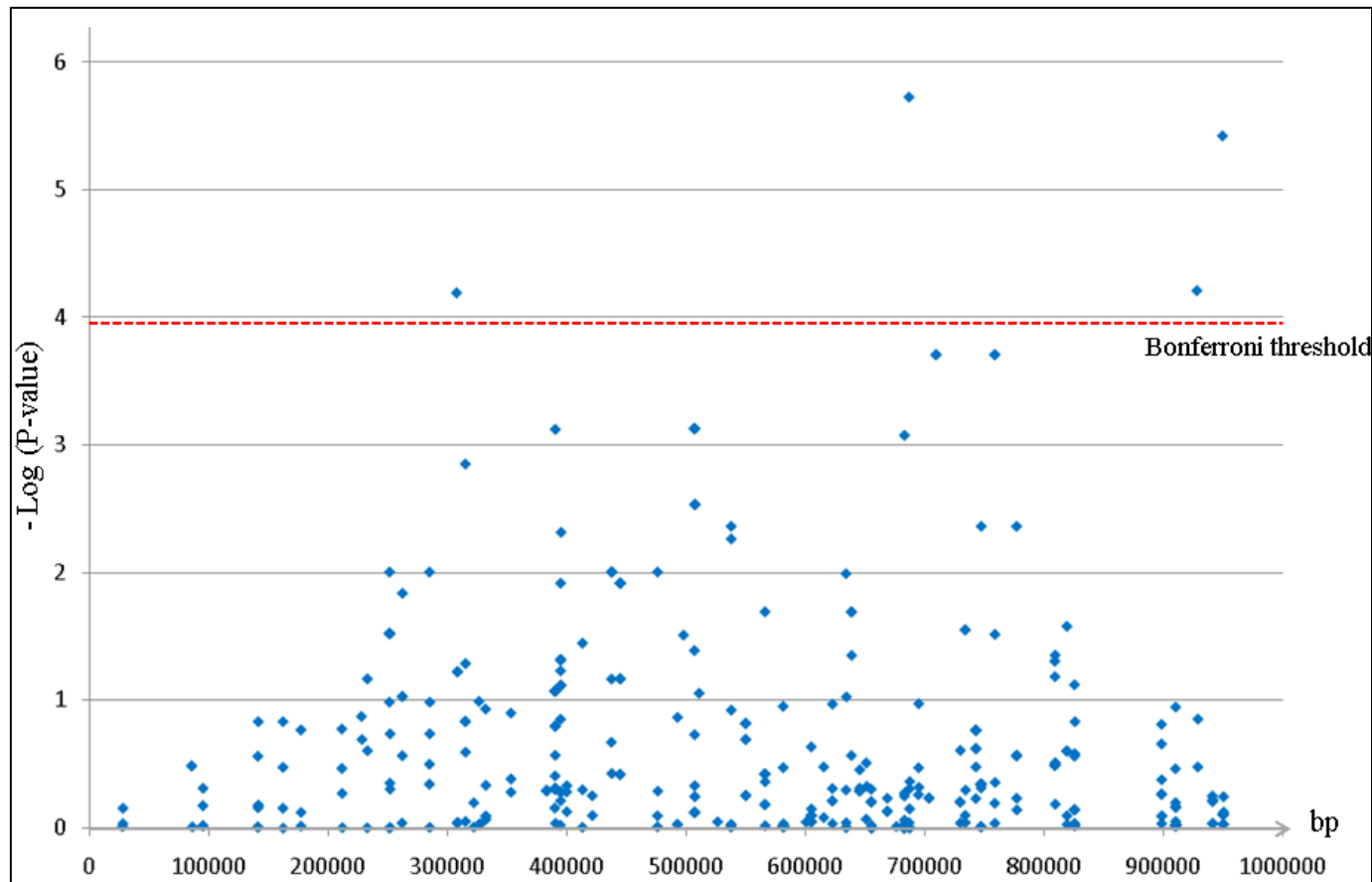


file 5], whereas only one fragment was monomorphic in the domesticated compartment (VV05795A). This was not the case for the 77 random fragments [additional file 3]. In addition, we found that the wine cultivar Orbois, like Ugni Blanc, was completely homozygous specifically in the *Flb* region (data not shown).

Under the hypothesis that *Flb* was one of the genes under selection during grape domestication, we expected to find traces of selection in the cultivated compartment associated with a difference of nucleotide diversity between the cultivated and wild compartments. Eight sequenced gene fragments in the *Flb* region were particularly interesting because they showed such possible traces of selection in the cultivated pool of genotypes (previous paragraph; Table 3). Four out of the eight gene fragments showed differences in nucleotide diversity between the two compartments (VV05791A, VVC2897A, VVC2901A and VVC2892A; Table 3). However, the wild *V. vinifera* sample showing over all the genome a lower diversity than the cultivated *V. vinifera* sample, we could conclude to a significant nucleotide diversity difference between wild and cultivated compartment only in the case where there was a decreasing of nucleotide diversity in the cultivated sample in comparison to the wild sample. Only one out of eight gene fragments (VVC2897A) showed such significant higher nucleotide diversity ( $\pi$ ) in the wild compartment compared to the cultivated compartment. This gene encodes a putative glyceraldehyde-3-phospho-dehydrogenase (Table 3). VVC2897A was re-sequenced in Ugni Blanc and Ugni Blanc mutant, showing no polymorphism in the part of the coding region they contained (data not shown).

### ***Flb* region showed significant LD and a possible association with berry size variation**

In order to check if there was linkage disequilibrium (LD) between the genes possibly under selection, LD was evaluated along the entire *Flb* region. Two sub-regions were highlighted [additional file 7]. The first one, close to the telomere, contained two out of the eight genes possibly under selection (VVC2981A and VVC2946A), showed lower nucleotide diversity (Figure 5) and several gene fragments with no SNP in the wild pool. Moreover, in this sub-region, few significant LD was observed between the different gene fragments in both cultivated and wild pools [additional file 7]. The second sub-region contained six out of eight genes possibly under selection and showed high nucleotide diversity and a significant LD between and within some gene fragments in the cultivated and wild pools (Figures 5 and 6). Most of the SNP found in the four out of the six genes possibly under selection showed intragenic LD, in the cultivated pool, and for two of them (VVC2901A and VVC2885A), an



**Figure 7.** Association tests for berry weight. Level of association between SNP markers and the average berry weight along the *Flb* region, in the cultivated *V. vinifera* sample. The Bonferroni threshold is equal to 1.12E-4.

**Table 5.** SNP significantly associated with the variation of berry weight in the cultivated and wild pool of *V. vinifera*.

Fragment	Position of the SNP	P-value <sup>\$</sup>	Feature	SNP type	SNP frequency	Putative function
VVC2966A	307,938	6.44E-05	CDS	[A/T]	5.00%	Unknown
VV05786A	687,166	1.87E-06	CDS	[A/G]*	9.00%	Protein kinase
VV05777A	928,516	6.18E-05	Intron	[T/G]	8.00%	Catalase
VV05775A	949,972	3.78E-06	Intron	[A/T]	9.00%	Ribosome biogenesis protein

<sup>\$</sup>Significant P-value after Bonferroni correction (for  $\alpha=0.05$ )

\* Non-synonymous mutation

intergenic LD was found and extended with the adjacent gene fragment VV05782A (Figure 6). In the wild pool, only five out of the six gene fragments possibly under selection were polymorphic and could be used for the estimation of the LD in the second sub-region. Three of them showed intragenic and (excepted for VVC2897A) intergenic LD, together and with the gene fragment VV05782A as for the cultivated pool. Finally, the only gene possibly under selection showing significant nucleotide diversity difference between the two pools (VVC2897A) showed strong intragenic LD and with four adjacent gene fragments (VV05786A, VVC2907A, VV05785A and VVC2903A).

We searched for associations in the set of cultivated genotypes between the average weight of mature berries and the 447 out of 554 SNP from the *Flb* region with an allelic frequency >0.05. Such significant associations (Figure 7; [additional file 8]) were detected for four SNP in four gene fragments listed in the Table 5. None of them corresponded to the genes showing a significant deviation from neutrality of the Tajima's D parameter. However, a significant association was found with a non synonymous SNP from a gene fragment (VV05786A; Table 5) showing LD with the only gene fragment possibly under selection with a high nucleotide diversity in the wild pool than in the cultivated pool, VVC2897A.

## **Discussion**

With an initial objective to develop markers tightly flanking the *Flb* mutation, 62 genomic regions were scanned for polymorphism along a 1.4 Mb region at the top of chromosome 18, where the mutation was previously located [21]. These regions were either genotyped or sequenced in the genotype carrying the mutation, Ugni Blanc mutant, its wild type (Ugni Blanc) and Chardonnay, which was the other parent of a full sib family segregating for the mutation. The sequenced fragments or markers were completely homozygous in Ugni Blanc and Ugni Blanc mutant, with one marker analyzed each 23 kb in average. Indeed, while analyzing the genome sequence of the heterozygous grapevine cultivar Pinot Noir, Velasco et al [43] showed that, like in other heterozygous species, the frequency of SNP or INDEL varied along the grapevine genome and found some evidence for scarce quasi-homozygous areas. Here we describe a region of 1 Mb probably completely homozygous that raised two questions. First, as Velasco et al [43] showed that over 65 Mb of sequence are hemizygous in Pinot Noir, we wanted to check if our observations were due to a real homozygosity or to a deletion of a large portion of the top of chromosome 18 in one haplotype of Ugni Blanc. We addressed this issue by two different experiments (Figure 2), a qPCR estimation of the

number of copies of a single gene in the homozygous area (*FL*) compared to genes elsewhere in the genome (three *HMGCoA* genes). The same number of copies was estimated for *FL* gene for Ugni Blanc mutant and Chardonnay which is heterozygous in this region. Second, a BAC-FISH hybridization on Ugni Blanc mutant metaphase chromosomes using a BAC clone located in the area was carried out and showed a signal on both homologous chromosomes. We therefore un-ambiguously demonstrated that our observations corresponded to a real homozygosity in Ugni Blanc mutant. This would be consistent with the fact that hemizygous regions identified by Velasco et al [43] would mainly correspond to stretches of repeated sequences, which is not the case of the *Flb* region. These results raised the question whether this high level of homozygosity in Ugni Blanc mutant was restricted to the top of chromosome 18. The scoring of 480 SNP [44] and 20 SSR ([45], V. Laucou personal communication) regularly spread along the genome showed that whereas this cultivar seems slightly more homozygous in average than for instance Cabernet Sauvignon, Syrah or Chardonnay, the near complete homozygosity observed in the *Flb* region in Ugni Blanc mutant is not the rule on the rest of the genome and may be restricted to this region only. A mechanism which could explain the formation of such large homozygous region in a highly heterozygous out-crosser like grapevine would involve the repair of a DNA double-strand break [46]. When analyzing diversity in the cultivated germplasm, we observed that the cultivar Orbois was also completely homozygous for all fragments re-sequenced at the top of chromosome 18, and confirmed by qPCR assay that it was also due to real homozygosity (data not shown). Whatever its origin, this unexpected result made impossible the fine mapping of the mutation in the available segregating F1 population, which would necessitate the development of a F2 population.

Before having such a population available, we tested another possibility for reducing the interval carrying the *Flb* gene, based on the fact that *Flb* could be a gene selected during grape domestication. Indeed, the berry and seed phenotypes of Ugni Blanc mutant look like the phenotypes of wild *V. vinifera* seeds and berries [20,22]. We searched for signatures of selection in the *Flb* region in a set of cultivated genotypes. For this purpose, we sequenced in 33 individuals (26 cultivated and 7 wild genotypes) (i) 69 gene fragments for a total of 34,355 kb along 948 kb in the *Flb* region and (ii) 77 additional, totaling 48,874 kb spread along the 19 grapevine chromosomes. As already observed by Vezzulli et al [41], the nucleotide diversity was lower in average in the set of wild genotypes ( $\pi=0.27$ ) than in the set of

cultivated genotypes ( $\pi=0.35$ ). This difference was increased in the *Flb* region probably by the fact that a quarter of the fragments showed no sequence polymorphism at all among the wild genotypes whereas only one was in this case in the domesticated compartment. In the present work (but not in Vezzulli et al [42]) the sample of domesticated genotypes was selected after a comprehensive analysis of the world-wide largest collection of domesticated grapevine accessions [23] with the aim to retain a maximum of diversity. Unlike in Vezzulli et al [42], very few specific SNP were found in the set of wild genotypes (six SNP out of 554). All these observations, opposite to what has been observed in many other species [47,48], could be due to the fact that only seven to ten wild genotypes were sequenced in both studies and that their choice could not be driven by a maximization of the diversity along their complete natural area of growth. However, several surveys including accessions from the wild *V. vinifera* germplasm also showed this overall lower genetic diversity compared to the cultivated germplasm [49,50,51]. Indeed, small population sizes [52,53] as well as dioecy [54] could explain the observed reduced diversity in the wild *V. vinifera* gene pool, while multiple domestication events [50] and a continuous breeding process the larger diversity in the cultivated gene pool. All these results including ours will have to be confirmed with a larger sample of wild *V. vinifera*, taking into account all the geographic area where it grows and its overall genetic diversity.

Putative signatures of selection (Tajima's D parameter significantly deviating from neutrality associated with differences between the cultivated and wild pools in sequence diversity) in the cultivated pool of genotypes and SNP showing a significant association with berry weight variation were found in the *Flb* region (Figure 5). Moreover, LD was found in this region, the genes under selection presenting intragenic LD and intergenic LD with nearby genes (Figure 6), which strengthened the hypothesis that they might be under selection [48]. Like in Fournier-Level et al [55], no LD was however observed between the six out of eight genes possibly under selection. It is well known that the LD extent varies between the organisms and genomic regions [56]. In grapevine, a recent preliminary genome wide study confirmed the low extent of LD in average (less than 3-10 kb), suggesting a large effective size in the grapevine population at the origin of the current domesticated pool [57]. However, in recent history, vegetative propagation and long intervals between generations may have reduced the impact of recombination, maintaining extensive linkage disequilibrium in some regions under selection [58].

The significant genetic associations found between berry weight variation and SNP in this region were not found in the fragments putatively under selection. It is still possible that one of these genes is involved into the berry weight variation and that the causative sequence polymorphism was not in the exon fragments sequenced. Indeed, trait variation is often due to sequence variation in regulatory regions (see for instance [9,11,55]) and such variant sites may be more tightly linked to a neighbor gene. Interestingly, the orthologs in *Arabidopsis* for the six out of eight gene with signature of selection (At1g76540 for VVC2946A, At1g76400 for VV05791A, At1g42970 for VVC2897A, At1g42540 for VV05785A, At5g43820 for VVC2901 and At1g20696 for VVC2892A) and for the two genes with SNP associated with berry weight variation (At5g24306 for VV05786A and At1g42440 for VV05775A), are all expressed in flowers and showed a peak of expression at the beginning or during the flowering ([urgv.evry.inra.fr/projects/FLAGdb++](http://urgv.evry.inra.fr/projects/FLAGdb++)). In tomato, the genes involved into fruit size variation have been shown to be expressed in very early stages of its development, starting at floral development [2,11]. Finally, only one gene fragment, VVC2897A, might be under purifying selection in the cultivated pool with a major haplotype in cultivated pool and a LD extended to a neighbor gene which present a SNP significantly associated with berry weight variation (Figure 5, Figure 6). This would fit with the hypothesis that, like in tomato, the selection of a new haplotype by humans would have ensured the transition from berries with little flesh in wild grapevines to berries with more flesh in cultivated grapevines [2,11]. The real involvement of this gene into berry size variation and in the domestication syndrome remains however to be proven.

## Conclusions

While searching for SNP markers in coupling with the *Fleshless berry* mutation, we observed the occurrence of a 1 Mb homozygous region, not associated with repetitive sequences, in the grapevine otherwise highly heterozygous genome. We demonstrated the feasibility to use BAC-FISH on the very small grapevine chromosomes and provided a specific probe for the identification of chromosome 18 on cytogenetic map. Using this method, we showed that the observed homozygosity was not due to a large deletion.

We then searched for signatures of domestication for berry weight along the *Flb* region by re-sequencing 69 gene fragments in 26 domesticated and seven wild *V. vinifera* genotypes. We found putative signatures of selection associated with significant differences in nucleotide diversity between the cultivated and the wild pool only in one gene (VVC2897A) and also

SNP significantly associated with berry weight variation in three other genes among which one is in DL with VVC2897A. The involvement of these four genes into berry weight variation in grapevine remains to be proved by further functional experiments. In addition, we detected 554 SNP along the *Flb* region. These polymorphisms could serve to develop a genotyping chip useful for a future fine mapping of the *Flb* gene in a F2 population and for the analysis of genetic diversity in larger sets of wild and cultivated genotypes.

## Authors' contributions

CH designed the primers for fragment re-sequencing, participated to their sequencing, analyzed the results, chose the BAC and prepared the root tips for the FISH, draft and corrected the paper. RB, M-CLP and DB were responsible for the sequencing of the fragments. RB and AC participated to the sequence analysis and made the RT-PCR experiment. JC and LT analyzed the SSR and genotyped SNP polymorphism. CG and AD helped with bio-informatics (scripts, database queries...). J-PP and RB generated the set of reference fragment sequences along the genome. PT, TL, and AND provided the DNA for the core-collection and plant phenotypes. M-SV and OC did the BAC-FISH experiments. A-FAB design the experiment, supervised it, drafted and corrected the manuscript. All authors read and approved the final manuscript.

## Acknowledgements

This study was supported by INRA, ANR and the French Ministry of Research. We thank Isabelle Le Clainche, Audrey Weber, Sylvain Santoni, Christophe Lepage and all the team of the Vassal germplasm collection for technical assistance; Stéphane Nicolas, Aurélie Siberchicot, Christine Cierco and Brigitte Mangin for providing a R routine for LD calculation and helpful discussions; Marie-Laure Martin Magniette for help in the statistical analysis, Dr Mark R. Thomas from the CSIRO Plant Industry, Australia for providing unpublished DNA marker information and Loïc Le Cunff for helpful discussions.

## References

- [1] **Champagnol F**: Critères de qualité de la vendange. In *Oenologie, Fondements Scientifiques et Technologiques*. Edited by Flanzy C. Paris: Lavoisier Tec &Do; 1998:653-659
- [2] **Tanksley SD**: The genetic, developmental, and molecular bases of fruit size and shape variation in tomato. *Plant Cell* 2004, 16:181-189
- [3] **Lozano R, Giménez E, Cara B, Capel J, Angosto T**: Genetic analysis of reproductive development in tomato. *Int J Dev Biol* 2009, 53:1635-1648
- [4] **Fridman E, Carrari F, Iiu Y-S, Fernie A, Zamir D**: Zooming in on a quantitative trait for tomato yield using interspecific introgressions. *Science* 2004, 305:1786-1789
- [5] **Causse M, Duffe P, Gomez M-C, Buret M, Damidaux R, Zamir D, Gur A, Chevalier C, Lemaire-Chamley M, Rothan C**: A genetic map of candidate genes and QTL involved in tomato fruit size and composition. *J Exp Bot* 2004, 55:1671-1685



- [6] **Lippman ZB, Semel Y, Zamir D**: An integrated view of quantitative trait variation using tomato interspecific introgression lines. *Curr Opin Genet & Dev* 2007, 17:545-552
- [7] **Chaïb J, Lecomte L, Buret M, Causse M**: Stability over genetic backgrounds, generations and years of quantitative trait locus (QTL) for organoleptic quality in tomato. *Theor Appl Genet* 2006, 112:934-944
- [8] **Menda N, Semel Y, Peled D, Eshed Y, Zamir D**: In silico screening of a saturated mutation library of tomato. *The Plant J* 2004, 38:861-872
- [9] **Frary A, Nesbitt TC, Frary A, Grandillo S, van der Knaap E, Cong B, Liu J, Meller J, Elber R, Alpert KB, Tanksley SD**: fw2.2: A Quantitative Trait Locus key to the evolution of tomato fruit size. *Science* 2000, 289:85-88
- [10] **Bertin N, Causse M, Brunel B, Tricon D, Genard M**: Identification of growth processes involved in QTL for tomato fruit size and composition. *J Exp Bot* 2009, 60:237-248
- [11] **Cong B, Barrero LS, Tanksley SD**: Regulatory change in YABBY-like transcription factor led to evolution of extreme fruit size during tomato domestication. *Nat Genet* 2008, 40:800-804
- [12] **Cheniclet C, Rong WY, Causse M, Frangne N, Bolling L, Carde J-P, Renaudin J-P**: Cell expansion and endoreduplication show a large genetic variability in pericarp and contribute strongly to tomato fruit growth. *Plant Physiol* 2005, 139:1984-1994
- [13] **This P, Lacombe T, Thomas MR**: Historical origins and genetic diversity of wine grapes. *Trends Genet* 2006, 22:511-519
- [14] **Boursiquot JM, Dessup M, Rennes C**: Distribution des principaux caractères phénologiques, agronomiques et technologiques chez *Vitis vinifera* L. *Vitis* 1995, 34:31-35
- [15] **Ollat N, Diakou-Verdin P, Carde J-P, Barrieu F, Gaudillère J-P, Moing A**: Grape berry development: a review. *J Int Sci Vigne Vin* 2002, 36:109-131
- [16] **Doligez A, Bouquet A, Danglot Y, Lahogue F, Riaz S, Meredith P, Edwards J, This P**: Genetic mapping of grapevine (*Vitis vinifera* L.) applied to the detection of QTL for seedlessness and berry weight. *Theor Appl Genet* 2002, 105:780-795
- [17] **Cabezas JA, Cervera MT, Ruiz-Garcia L, Carreno J, Martinez-Zapater JM**: A genetic analysis of seed and berry weight in grapevine. *Genome* 2006, 49:1572-1585
- [18] **Mejia N, Gebauer M, Muñoz L, Hewstone N, Muñoz C, Hinrichsen P**: Identification of QTL for seedlessness, berry size, and ripening date in a seedless x seedless progeny. *Am J Enol Vitic* 2007, 58:499-507
- [19] **Costantini L, Battilana J, Lamaj F, Fanizza G, Grando M**: Berry and phenology-related traits in grapevine (*Vitis vinifera* L.) From Quantitative Trait Loci to underlying genes. *BMC Plant Biology* 2008, 17:8-38

- [20] **Fernandez L, Romieu C, Moing A, Bouquet A, Maucourt M, Thomas MR, Torregrosa L**: The grapevine *Fleshless berry* mutation. A unique genotype to investigate differences between fleshy and nonfleshy fruit. *Plant Physiol* 2006, 140:537-547
- [21] **Fernandez L, Doligez A, Lopez G, Thomas MR, Bouquet A, Torregrosa L**: Somatic chimerism, genetic inheritance, and mapping of the *Fleshless berry* (*Flb*) mutation in grapevine (*Vitis vinifera* L.). *Genome* 2006, 49:721-728
- [22] **Terral J-F, Tabard E, Bouby L, Ivorra S, Pastor T, Figueiral I, Picq S, Chevance J-B, Jung C, Fabre L, Tardy C, Compan M, Bacilieri R, Lacombe T, This P**: Evolution and history of grapevine (*Vitis vinifera*) under domestication: new morphometric perspectives to understand seed domestication syndrome and reveal origins of ancient European cultivars. *Annals of Botany* 2010, 105:443-55
- [23] **Le Cunff L, Fournier-Level A, Laucou V, Vezzulli S, Lacombe T, Adam-Blondon A-F, Boursiquot J-M, This P**: Construction of nested core collections to optimize the exploitation of natural diversity in *Vitis vinifera* L. subsp. *sativa*. *BMC Plant Biology* 2008, 2:8-31
- [24] **Jaillon O, Aury J-M, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C, Vezzi A, Legeai F, Hugueney P, Dasilva C, Horner D, Mica E, Jublot D, Poulain J, Bruyere C, Billault A, Segurens B, Gouyvenoux M, Ugarte E, Cattonaro F, Anthouard V, Vico V, Del Fabbro C, Alaux M, Di Gaspero G, Dumas V, Felice N, Paillard S, Juman I, Moroldo M, Scalabrin S, Canaguier A, Le Clainche I, Malacrida G, Durand E, Pesole G, Laucou V, Chatelet P, Merdinoglu D, Delledonne M, Pezzotti M, Lecharny A, Scarpelli C, Artiguenave F, Pé E, Valle G, Morgante M, Caboche M, Adam-Blondon A-F, Weissenbach J, Quétier F, Wincker P**: The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 2007, 449:463-468
- [25] **Rozen S, Skaletsky HJ**: Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* 2000, 132:365-86
- [26] **Philippe R, Courtois B, McNally KL, Mournet P, El-Malki R, Le Paslier MC, Fabre D, Billot C, Brunel D, Glaszmann JC, This D**: Structure, allelic diversity and selection of *Asr* genes, candidate for drought tolerance, in *Oryza sativa* L. and wild relatives. *Theor Appl Genet* 2010, 121:769-787
- [27] **Salmaso M, Malacarne G, Troggio M, Faes G, Stefanini M, Grando MS, Velasco R**: A grapevine (*Vitis vinifera* L.) genetic map integrating the position of 139 expressed genes. *Theor Appl Genet* 2008, 116:1129-1143
- [28] **Thomas MR, Cain P, Scott NS**: DNA Typing of Grapevines. A universal methodology and database for describing cultivars and evaluating genetic relatedness. *Plant Mol Biol* 1994, 25:939-949
- [29] **Reid KE, Olsson N, Schlosser J, Peng F, Lund ST**: An optimized grapevine RNA isolation procedure and statistical determination of reference genes for real-time RT-PCR during berry development. *BMC Plant Biology* 2006, 14:6-27

- [30] **Leflon M, Eber F, Letanneur JC, Chelysheva L, Coriton O, Huteau V, Ryder CD, Barker G, Jenczewski E, Chevre AM**: Pairing and recombination at meiosis of *Brassica rapa* (AA) x *Brassica napus* (AACC) hybrids. *Theor Appl Genet* 2006, 113:1467-1480
- [31] **Gerlach W, Bedbrook J**: Cloning and characterization of ribosomal RNA genes from wheat and barley. *Nucl Acids Res* 1979, 7:1869-1885
- [32] **Takahashi M, Matsuda F, Margetic N, Lathrop M**: Automated identification of single nucleotide polymorphisms from sequencing data. *Journal of Bioinformatics and Computational Biology* 2003, 1:253-265
- [33] **Staden, R**: The Staden Sequence Analysis Package. *Molecular Biotechnology* 1996, 5:233-241
- [34] **Tajima F**: Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 1989, 123:585-595
- [35] **Roger AR, Huff C**: Linkage disequilibrium between loci with unknown phase. *Genetics* 2009, 182:839-844
- [36] **Pritchard JK, Stephens M, Donnelly P**: Inference of population structure using multilocus genotype data. *Genetics* 2000, 155: 945-959
- [37] **Adam-Blondon AF, Roux C, Claux D, Butterlin G, Merdinoglu D, This P**: Mapping 245 SSR markers on the *Vitis vinifera* genome: a tool for grape genetics. *Theor Appl Genet* 2004, 109:1017-1027
- [38] **Di Gaspero G, Cipriani G, Adam-Blondon AF, Testolin R**: Linkage maps of grapevine displaying the chromosomal locations of 420 microsatellite markers and 82 markers for R-gene candidates. *Theor Appl Genet* 2007, 114:1249-1263
- [39] **Lamoureux D, Bernole A, Le Clainche I, Tual S, Thareau V, Paillard S, Legeai F, Dossat C, Wincker P, Oswald M, Merdinoglu D, Vignault C, Delrot S, Caboche M, Chalhoub B, Adam-Blondon AF**: Anchoring of a large set of markers onto a BAC library for the development of a draft physical map of the grapevine genome. *Theor Appl Genet* 2006, 113: 344-356
- [40] **Riaz S, Dangl GS, Edwards KJ, Meredith CP**: A microsatellite marker based framework linkage map of *Vitis vinifera* L. *Theor Appl Genet* 2004, 108:864-872
- [41] **Troggio M, Malacarne G, Coppola G, Segala C, Cartwright DA, Pindo M, Stefanini M, Mank R, Moroldo M, Morgante M, Grando MS, Velasco R**: A dense single-nucleotide polymorphism-based genetic linkage map of grapevine (*Vitis vinifera* L.). Anchoring Pinot Noir Bacterial Artificial Chromosome Contigs. *Genetics* 2007, 176:2637-2650
- [42] **Vezzulli S, Micheletti D, Riaz S, Pindo M, Viola R, This P, Walker MA, Troggio M, Velasco R**: A SNP transferability survey within the genus *Vitis*. *BMC Plant Biology* 2008, 16:8-128

- [43] Velasco R, Zharkikh A, Troggio M, Cartwright DA, Cestaro A, Pruss D, Pindo M, Fitzgerald LM, Vezzulli S, Reid J, Malacarne G, Iliev D, Coppola G, Wardell B, Micheletti D, Macalma T, Facci M, Mitchell JT, Perazzolli M, Eldredge G, Gatto P, Oyzerski R, Moretto M, Gutin N, Stefanini M, Chen Y, Segala C, Davenport C, Demattè L, Mraz A, Battilana J, Stormo K, Costa F, Tao Q, Si-Ammour A, Harkins T, Lackey A, Perbost C, Taillon B, Stella A, Solovyev V, Fawcett JA, Sterck L, Vandepoele K, Grando SM, Toppo S, Moser C, Lanchbury J, Bogden R, Skolnick M, Sgaramella V, Bhatnagar SK, Fontana P, Gutin A, Van de Peer Y, Salamini F, Viola R: A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. *PLoS ONE* 2007, 2:e1326
- [44] Chaïb J, Torregrosa L, Mackenzie D, Corena P, Bouquet A, Thomas MR: The grape microvine – a model system for rapid forward and reverse genetics of grapevines. *The Plant J* 2010, 62:1083-1092
- [45] Di Vecchi-Staraz M, Bandinelli R, Boselli M, This P, Boursiquot JM, Laucou V, Lacombe T, Vares D: Genetic structuring and parentage analysis for evolutionary studies in grapevine: Kin group and origin of the cultivar Sangiovese revealed. *Journal of the American Society for Horticultural Science* 2007, 132:514-524
- [46] Puchta H: The repair of double-strand breaks in plants: mechanisms and consequences for genome evolution. *J Exp Bot* 2005, 56:1-14
- [47] Miller JC, Tankley SD: RFLP analysis of phylogenetic relationships and genetic variation in the genus *Lycopersicon*. *Theor Appl Genet* 1990, 80:437-448
- [48] Palaisa KA, Morgante M, Williams M, Rafalski A: Long-range patterns of diversity and linkage disequilibrium surrounding the maize Y1 gene are indicative of an asymmetric selective sweep. *PNAS* 2004, 101:9885-9890
- [49] Aradhya MK, Dangl GS, Prins BH, Boursiquot J-M, Walker MA, Meredith CP, Simon CJ: Genetic structure and differentiation in cultivated grape, *Vitis vinifera* L. *Genet Res Camb* 2003, 81:179-192
- [50] Arroyo-Garcia R, Ruiz-Garcia L, Bolling L, Ocete R, Lopez MA, Arnold C, Ergul A, Söylemezoglu G, Uzun HJ, Cabello F, Ibanez J, Aradhya MK, Atanassov I, Balint S, Cenis JL, Costantini L, Gorislavets S, Grando MS, Klein BY, Govern PEM, Merdinoglu D, Pejic I, Pelsy F, Primikirios N, Risovannaya V, Roubelakis-Angelakis KA, Snoussi H, Sotiri P, Tamhankar S, This P, Troshin L, Malpica JM, Lefort F, Martinez-Zapater JM: Multiple origins of cultivated grapevine (*Vitis vinifera* L. ssp. *sativa*) based on chloroplast DNA polymorphisms. *Molecular Ecology* 2006, 15:3707-3714
- [51] Grassi F, De Mattia F, Zecca G, Sala F, Labra M: Historical isolation and Quaternary range expansion of divergent lineages in wild grapevine. *Biological Journal of the Linnean Society* 2008, 95:611-619
- [52] Levadoux L: Les Populations sauvages et cultivées de *Vitis vinifera* L. *Annales de l'Amélioration de Plantes* 1956, 6:59-118

- [53] **Arnold C, Gillet F, Gobat JM**: Occurrence of the wild vine *Vitis vinifera* ssp. *silvestris* in Europe. *Vitis* 1998, 37:159-170
- [54] **Di Vecchi-Staraz M, Laucou V, Bruno G, Lacombe T, Gerber S, Bourse T, Boselli M, This P**: Low level of pollen-mediated gene flow from cultivated to wild grapevine: consequences for the evolution of the endangered subspecies *Vitis vinifera* L. subsp. *silvestris*. *J Heredity* 2009, 100:66-75
- [55] **Fournier-Level A, LeCunff L, Gomez C, Doligez A, Ageorges A, Roux C, Bertrand Y, Souquet JM, Cheynier V This P**: Quantitative genetic bases of anthocyanin variation in grape (*Vitis vinifera* L. ssp. *sativa*) berry: a quantitative trait locus to quantitative trait nucleotide integrated study. *Genetics* 2009, 183:1127-1139
- [56] **Gupta P.K, Rustgir S, Kulwal PL**: Linkage disequilibrium and association studies in higher plants: Present status and future prospects. *Plant Molecular Biology* 2005, 57:461-485
- [57] **Myles S, Chia J-M, Hurwitz B, Simon C, Zhong GY, Buckler E, Ware D**: Rapid genomic characterization of the genus *Vitis*. *Plos One* 2010, 5:e8219
- [58] **Barnaud A, Lacombe T and Doligez A**: Linkage disequilibrium in cultivated grapevine, *Vitis vinifera* L. *Theor Appl Genet* 2009, 112:708-716



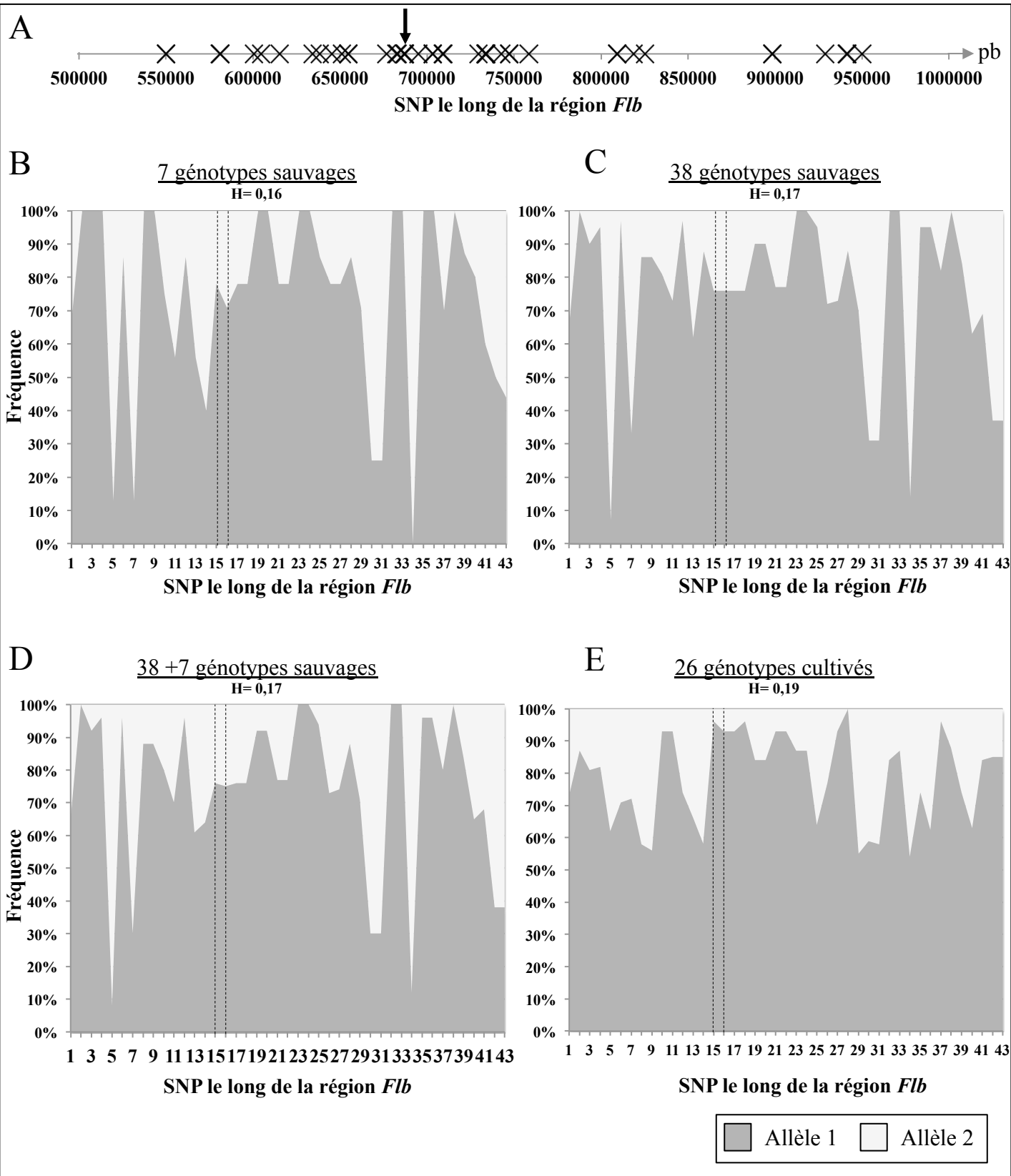
## 5.3. Analyses complémentaires

J'ai pu mettre en évidence une putative glyceraldéhyde-3-phospho-dehydrogenase (fragment de séquence VVC2897A) qui présente une déviation significative du D de Tajima, par rapport à l'hypothèse de neutralité, avec une valeur située en queue de la distribution du D de Tajima calculée sur des gènes répartis aléatoirement sur le génome, suggérant que ce gène a été soumis à la sélection. Ce gène présente également une diversité nucléotidique plus faible dans l'échantillon de vignes cultivées que dans celui de vignes sauvages, ce qui est l'inverse de ce qui est observé sur les fragments re-séquencés aléatoirement le long du génome. Cependant, il faut noter que l'échantillon de vignes sauvages utilisé dans cette étude ne comprend que sept génotypes présentant une diversité génétique plus faible que celle des génotypes cultivés, ce qui est à l'opposé de ce qui est observé en général chez d'autres espèces. De plus, ce gène n'a pas été trouvé directement associé à la variation du poids de la baie dans l'étude préliminaire sur le panel de 26 vignes cultivées. Quatre SNP dans la région *Flb* présents dans quatre autres fragments de gènes (VV05786A, VVC2966A, VV05777A et VV05775A) ont, eux, été trouvés significativement associés à la variation du poids de la baie dans le compartiment cultivé dans cette même étude mais ne présentent pas de trace de sélection.

Afin de confirmer l'implication de ces gènes dans le processus de développement de la pulpe et leur rôle possible dans le processus de domestication: (1) la diversité de séquence de la région *Flb* a été analysée dans un set plus large de vignes sauvages, (2) un test d'association en relation avec la variation du poids de la baie a été réalisé dans un panel plus important de 240 individus peu apparentés, structurés en trois sous-populations et maximisant la diversité génétique de la vigne cultivée, et (3) l'expression de ces gènes candidats a été quantifiée au cours du développement de la fleur et de la baie pour des génotypes produisant de petites ou de grosses baies.

### 5.3.1. Diversité génétique au locus *Flb* chez *V. vinifera* ssp *sylvestris*

Un panel SNPlex de 47 SNP dans 28 gènes répartis le long de la région *Flb* de part et d'autre du gène VVC2897A a été développé (1 à 4 SNP ou INDEL par gène avec une fréquence >5% dans l'échantillon; Annexe 4-A). Ce panel de SNP a été utilisé pour génotyper un ensemble de 38 génotypes *V. vinifera* sauvages supplémentaires non apparentés, de provenances géographiques diverses et maximisant la diversité génétique de la vigne sauvage



**Figure 5.2:** Répartition et fréquence allélique des SNP génotypés dans les différents échantillons. **(A)** Répartition des SNP le long de la région *Flb*. Fréquence allélique des SNP chez **(B)** les 7 sauvages de l'échantillon de re-séquençage, **(C)** et les 38 sauvages supplémentaires, **(D)** l'ensemble des sauvages et **(E)** les 26 génotypes cultivés. L'allèle le plus fréquent dans le set de vigne cultivées correspond à l'allèle 1. H: taux d'hétérozygotie moyen observé sur la région *Flb*. La flèche noire et les traits pointillés correspondent aux SNP du fragment VVC2897A.

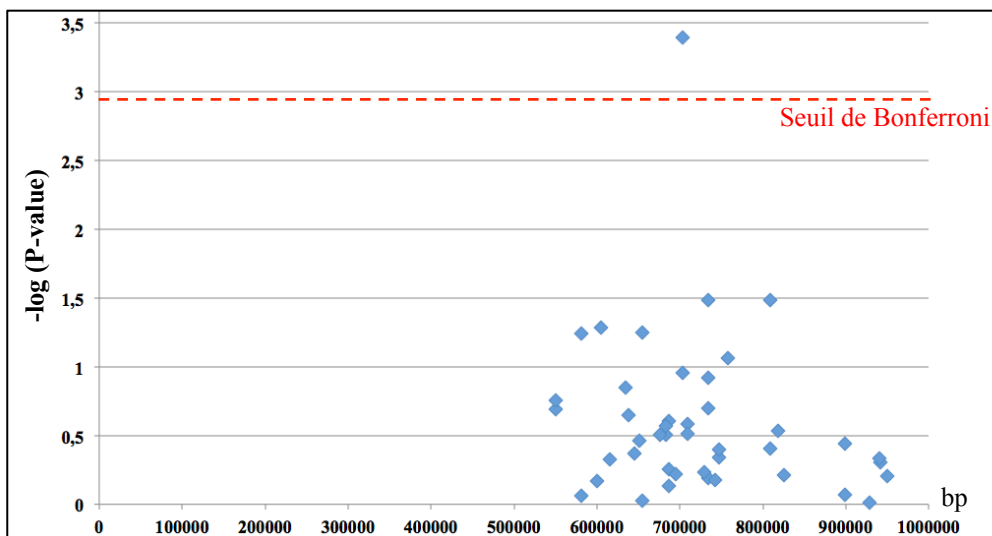


présente sur le Domaine de Vassal (Annexe 4-B). Les 26 génotypes cultivés ainsi que les 7 génotypes sauvages précédemment séquencés ont été génotypés comme témoins positifs de l'expérience. Quatre des 47 SNP ont été éliminés pour diverses raisons: contamination du témoin négatif de l'expérience, absence de signal (pas d'hybridation de l'amorce) ou profil en « patte d'oie » (SNP triallélique ou amorce multilocus). Au final 43 SNP ont donc été analysés (annexe 4-A).

Il n'était pas possible de calculer la diversité nucléotidique ( $\Pi$ ) à partir des données de génotypage. La diversité génétique de la région *Flb* a donc été analysée à travers deux paramètres: le taux moyen d'hétérozygotie observée ( $H$ ) et la fréquence des allèles à chaque SNP (Figure 5.2-B et C).

Le profil de la fréquence des allèles des 43 SNP répartis le long de la région *Flb*, est quasiment identique dans le groupe des sept sauvages précédemment étudiés et dans le groupe des 38 sauvages supplémentaires (Figure 5.2-A). De plus, le taux moyen d'hétérozygotie de ces deux panels de génotypes est très proche respectivement 0,16 et 0,17. On peut également remarquer que six des 43 SNP génotypés ne sont pas polymorphes chez les sauvages (l'allèle 1 a une fréquence égale à 100%). Par contre, huit autres SNP qui étaient monomorphes dans le groupe de sept sauvages se sont avérés polymorphes dans celui de 38, mais avec une fréquence du second allèle très faible. Dans l'ensemble, le panel des sept sauvages semble donc être représentatif de celui des 38 *V. vinifera* sauvages. Ces génotypes n'étant pas apparentés et provenant de différentes régions géographiques, on peut supposer qu'ils sont également relativement représentatifs de la vigne sauvage, et que les résultats obtenus précédemment sont confortés.

Deux des 43 SNP génotypés se trouvent dans le fragment de séquence VVC2897A (SNP numéro 15 et 16; Figure 5.2-C). Ces SNP présentent une diversité génétique plus faible dans le groupe de vignes cultivées que dans le groupe de vignes sauvages avec un taux d'hétérozygotie respectivement de 0,04 et 0,08 pour les cultivées et 0,29 et 0,31 pour les sauvages, confirmant les résultats précédemment obtenus. Cette diminution de la diversité génétique chez les génotypes cultivés par rapport aux sauvages s'étend aux deux SNP suivants (SNP numéro 17 et 18; Figure 5.2-C) avec un taux d'hétérozygotie respectivement de 0,08 pour les cultivées et 0,29 pour les sauvages pour les deux SNP. Ces SNP se trouvent dans le fragment de séquence VV05786A qui était en déséquilibre de liaison avec le fragment VVC2897A dans l'étude précédente. De plus, ce fragment est également associé à la variation



**Figure 5.3:** Test d'association pour la variation du poids de la baie. Niveau d'association entre les 43 marqueurs présents dans la région *Flb* et le poids moyen de la baie dans un échantillon de 240 génotypes cultivés. Le seuil de Bonferonni est égal à 0.001.

**Tableau 5.1:** Critères à l'origine du choix des trois gènes pour l'étude transcriptomique.

Fragment de gène	D de Tajima significatif	Différence de diversité sauvage/cultivé	Association poids de la baie
VVC2897A	oui	oui	non
VV05786A	non	non	oui (3 sous-populations)
VV05785A	oui	non	oui (sous-échantillon SNPGrpMap)

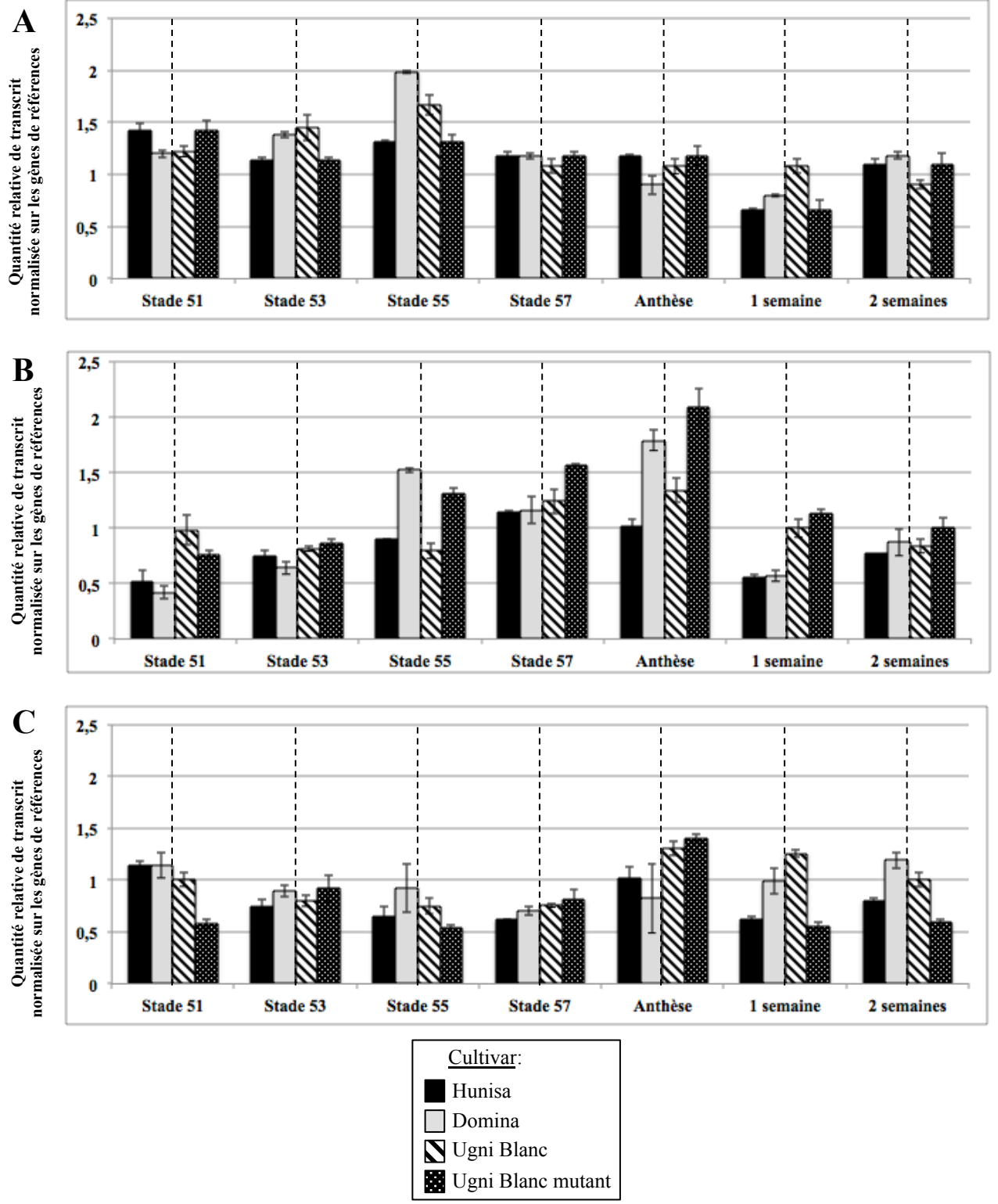
du poids de la baie dans l'étude préliminaire sur l'échantillon de 26 génotypes cultivés. Nous avons donc voulu confirmer ces associations entre les marqueurs SNP et la variation de la taille de la baie dans un échantillon plus important de génotypes cultivés.

### 5.3.2. Génétique d'association dans les 3 sous-populations

La recherche préliminaire d'associations entre les variations du poids de la baie et les variations nucléotidiques dans le panel de 26 génotypes cultivés avait permis de mettre en évidence une association avec quatre SNP dans les fragments de séquences VV05786A, VVC2966A, VV05777A et VV05775A.

Afin de confirmer ces résultats, les 43 SNP ont également été génotypés sur 278 génotypes cultivés non apparentés, structurés en 3 sous-populations et maximisant la variabilité génotypique de la vigne cultivée. Malheureusement, aucun des quatre SNP associés précédemment ne faisait partie du panel SNPlex en raison de contraintes de définition des amorces pour cette méthode de génotypage. Cependant 3 SNP sur le fragment VV05786A, 1 SNP sur VV05777A et 1 SNP sur VV05775A ont été génotypés. Les tests d'associations structurés entre les 43 SNP et le log du poids moyen de la baie à maturité ont été réalisés sur 240 des 278 génotypes du à des données phénotypiques manquantes, en tenant compte également, du lien d'apparentement de chaque individu (matrice d'apparentement identique à celle utilisée dans le chapitre 4) et l'erreur liée aux multi-tests effectués a été corrigée par la correction de Bonferroni pour un seuil à 0,05.

Aucune association significative n'a été trouvée entre la variation de la taille de la baie et les SNP présents dans le panel SNPlex correspondant aux quatre fragments de gènes précédemment identifiés. Une nouvelle association a été détectée entre le SNP 22 et la variation de la taille de la baie dans les 3 sous-populations (Figure 5.3). Ce SNP se trouve dans le fragment de séquence VV05785A, qui présente un D de Tajima significatif situé en queue de distribution du D de Tajima, calculé sur des fragments répartis aléatoirement sur le génome. Toutefois, ce fragment n'avait pas été retenu dans l'étude précédente, car il ne présentait pas de différence au niveau de la diversité nucléotidique entre les vignes sauvages et cultivées. Ce fragment se situe respectivement à 16 kb et 20 kb des fragments VV05786A et VVC2897A.



**Figure 5.4:** Quantité relative de transcrits correspondant aux gènes (A) VVC2897A, (B) VV05785A et (C) VV05786A, normalisée sur l'expression des gènes de référence au cours du développement de la fleur et de la baie, moyennée sur deux années. Les stades 51, 53, 55, 57, anthèse correspondent à la formation de la fleur (code international BBCH; Lorenz et al, 1994) et les stades 1 semaine et 2 semaines au nombre de semaines après l'anthèse donc aux stades précoces du développement de la baie. Les cultivars Hunisa, Domina, Ugni Banc et Ugni Blanc mutant ont respectivement de baies pesant en moyenne 6 g, 1 g, 2 g et 0.3 g.

Afin de mieux comprendre le rôle joué par ces gènes, une étude de leur expression (Tableau 5.1) dans des génotypes contrastés pour le poids de la baie a été réalisée au cours du développement de la fleur et de la baie.

### 5.3.3. Expression des gènes candidats au cours du développement de l'ovaire et de la baie

L'expression des gènes correspondant aux fragments de séquence VV05785A, VV05786A et VVC2897A, ainsi que celle de deux gènes de référence (une *aquaporine* et un facteur d'élongation: *EFL-α*; Annexe 4-C) ont été analysés par q-RT-PCR au cours du développement de la fleur (stade 51, 53, 55, 57 et à l'anthèse; stades correspondant au code international BBCH; Lorenz et al, 1994) et du développement précoce de la baie (1 et 2 semaines après l'anthèse) chez l'Ugni Blanc, l'Ugni Blanc mutant, le cv. Domina (produisant de petite baies <1g) et le cv. Hunisa (produisant de grosses baies >6g). La figure 5.4 présente la quantité relative de transcrits des différents gènes normalisée par celles des deux gènes de référence, pour les quatre génotypes à différents stades du développement. Nous avons privilégié les stades précoces du développement de l'ovaire et de la baie, en accord avec les résultats obtenus dans le chapitre 3, indiquant que la variation de la taille finale de la baie est déterminée précocement par le nombre de cellules de l'ovaire juste avant l'anthèse, mais également par le volume des cellules de la baie déterminé avant véraison. Les gènes impliqués dans ces processus s'expriment donc certainement dans les premières étapes de la formation de l'ovaire et/ou de la baie.

Les trois gènes s'expriment au cours du développement de l'ovaire et de la baie, avec un pic d'expression à l'anthèse (Figure 5.4) pour les gènes VV05785A et VV05786A et plus précocement (stade 55) pour le gène VVC2897A.

Le gène VVC2897A (qui présente des traces de sélection et une différence de diversité nucléotidique entre sauvages et cultivés; Tableau 5.1) et le gène VVC5785A (qui présente des traces de sélection et qui est associé à la variation du poids de la baie dans les 3 sous-populations; Tableau 5.1), présentent des différences d'expression entre les génotypes pour les stades du développement floral et également après l'anthèse. Cependant, le lien entre ces différences d'expression et les différences de poids de baies des génotypes n'est pas évident puisque le cv. Hunisa, produisant de très grosses baies (> 6 g), a souvent le même profil d'expression que l'Ugni Blanc mutant produisant des baies d'environ 0.3 g.

Enfin, le gène VV05786A (associé à la variation du poids de la baie dans le sous-échantillon SNPGrpMap; Tableau 5.1) présente des différences d'expression entre des génotypes contrastés pour la taille de la baie pour le stade 55 du développement floral et l'anthèse. A l'anthèse, l'expression de VV05786A est inversement proportionnelle à la taille de la baie. En effet, le *cv.* Hunisa produisant les plus grosses baies exprime faiblement le gène (quantité relative =1; Figure 5.4), tandis que pour les génotypes produisant de moyennes baies (*cv.* Ugni Blanc) et de petites baies (*cv.* Domina) l'expression du gène est de 1,3 et 1,7 respectivement, et > 2 chez l'Ugni Blanc mutant. Cette corrélation est moins nette au stade 55, puisque que les *cv.* Ugni Blanc et Hunisa expriment de la même façon le gène. Ces résultats suggèrent que le gène VV05786A pourrait effectivement jouer un rôle dans la variation de la taille de la baie *via* une différence d'expression de ce gène principalement à l'anthèse.

## **5.4. Conclusion du chapitre**

La combinaison des approches de génétique d'association dans deux panels d'individus de taille différente et de recherche de traces de sélection, fait ressortir trois gènes. Après avoir vérifié que la représentativité de notre panel d'individus sauvages était correcte, et complété les tests d'association, nous avons voulu valider nos résultats avec des études d'expression de gènes candidats qui ressortaient de notre étude. Celles-ci ont montré qu'ils étaient bien exprimés dans les fleurs et les baies avec un pic d'expression à l'anthèse ou juste avant. Cependant, l'analyse du transcriptome de fleurs et baies à différents stades du développement chez la vigne à l'aide d'une puce contenant 25 471 sondes a révélé que la moitié des gènes ciblés par cette puce s'exprimaient dans au moins un des stades (Zamboni et al, 2010). Il est donc possible que nous ayons échantillonné par hasard des gènes s'exprimant dans ces organes.

L'ensemble des résultats suggère que VVC2897A, qui présente des traces de sélection, ne serait finalement pas impliqué dans la variation de la taille de la baie puisqu'il ne ressort pas dans les études de génétique d'association et que l'étude de son profil d'expression au cours du développement de la fleur et de la baie n'est pas corrélé à la variation du poids de la baie entre les génotypes étudiés. VV05785A est bien associé à la variation de la taille de la baie, mais ce résultat n'a pas été confirmé par l'analyse d'expression du gène. Il est possible que nous soyons passés à côté du stade clé où une différence d'expression très importante

entre les génotypes pour ce gène détermine la différence du poids final de la baie, comme il est possible que les différences soient déterminées au niveau du fonctionnement de la protéine.

Enfin, un seul gène (VV08786A) présente des différences d'expression concordantes avec la variation de poids de baies entre l'Ugni Blanc et l'Ugni Blanc mutant, et entre un cultivar à très grosses baies et un autre à petites baies. Cependant, ce gène n'est pas associé à la variation de la taille de la baie dans les 3 sous-populations. L'ensemble de ces résultats ne permet pas de conclure de façon certaine quant à l'implication de ce gène dans la variation de la taille de la baie.

Il n'est pas évident, d'après nos résultats de génétique d'association et d'expression, de relier les traces de sélection à la variation du poids de la baie. En effet, n'ayant pu phénotyper l'intégralité de l'échantillon de génotypes sauvages pour le poids de la baie, nos tests d'associations ont été réalisés dans le compartiment cultivé et le gène *Flb* pourrait avoir été très fortement sélectionné pendant la domestication de la vigne, et n'être plus présent que sous la forme d'un allèle chez les génotypes cultivés, donc ne pas être associé avec la variation du poids de la baie à l'intérieur de ce compartiment. Une étude de génétique d'association, conduite dans les deux compartiments serait donc pertinente, mais se heurte au petit nombre de génotypes sauvages femelles produisant suffisamment de baies pour être génotypés. Une approche de détection de QTL dans des populations dérivées de croisements entre sauvages et cultivés serait finalement la plus intéressante à développer. Elle permettrait de voir si un QTL de poids de baie se positionne dans la région du gène *Flb*.

Les traces de sélection observées pour les gènes VVC2897A et VV05785A pourraient également être liées à d'autres caractères phénotypiques distinguant les génotypes sauvages des cultivés, comme la compaction de la grappe, le contenu en polyphénols, ou dues à la sélection naturelle liée à l'adaptation à l'environnement... Cette dernière hypothèse permettrait d'expliquer l'étendue du DL dans la région *Flb* observé chez les génotypes cultivés, mais également chez les sauvages. En effet, le déséquilibre de liaison (DL) est beaucoup plus important dans la région *Flb*, que ce qui a déjà été calculé chez la vigne sur le reste du génome (Lijavetzky et al, 2007; Myles et al, 2010). Enfin, il n'est pas à exclure que les résultats que nous avons obtenus puissent être en quelque sorte des « faux positifs » liés à notre échantillon de départ. L'ensemble de ces résultats illustre la difficulté des approches de détection de traces de sélection chez la vigne.

Le clonage positionnel du gène *Flb* permettrait sans aucun doute d'avancer de façon significative dans la connaissance des mécanismes importants dans le développement de la pulpe du raisin. Cependant, j'ai montré que cette identification par clonage positionnel est très ralentie par l'absence de polymorphisme dans la région du gène chez le génotype porteur de la mutation. Par ailleurs, la population Picovine x Ugni Blanc mutant s'avère également délicate à phénotyper pour ce caractère en raison d'une interaction entre le phénotype *Fleshless berry* et le phénotype *dwarf* et de l'instabilité de la mutation dans la descendance (Fernandez et al, 2006c; Chaïb et al, 2010). Cependant, mes travaux donnent accès à une ressource importante en polymorphisme dans la région qui devrait pouvoir être exploitée dans des descendance F2.

La région de 1Mb contenant le gène *Flb* semble avoir une évolution très particulière puisqu'elle peut se présenter sous une forme complètement homozygote chez la vigne qui est connue pour avoir un génome hautement hétérozygote et être sensible à la dépression de consanguinité. Dans le sous-échantillon de 32 génotypes *V. vinifera* que nous avons utilisé au cours de cette étude, non seulement l'Ugni Blanc, mais également l'Arbois, qui n'est pas apparenté avec l'Ugni Blanc (J-M. Boursiquot communication personnelle), sont homozygotes pour cette région. Dans notre échantillon de 278 génotypes cultivés supplémentaires, les 43 SNP génotypés étaient également complètement homozygotes chez les cultivars: Baresana, Ohanes, Ksvidinka, Tantovina, Garrido macho, Basicata, Fumin, Autumn seedless, Attiki, Gibert, Nehelescol, Totika, Urmidinka, Grain, Galotta, également non apparentés par construction de l'échantillon, laissant supposer que cette région pourrait également être homozygote chez ces cultivars. L'hypothèse que je propose est que cette homozygotie soit liée à un mécanisme de réparation de l'ADN après cassure, ce qui laisserait supposer une zone de fragilité qui favoriserait ce type d'évènement.

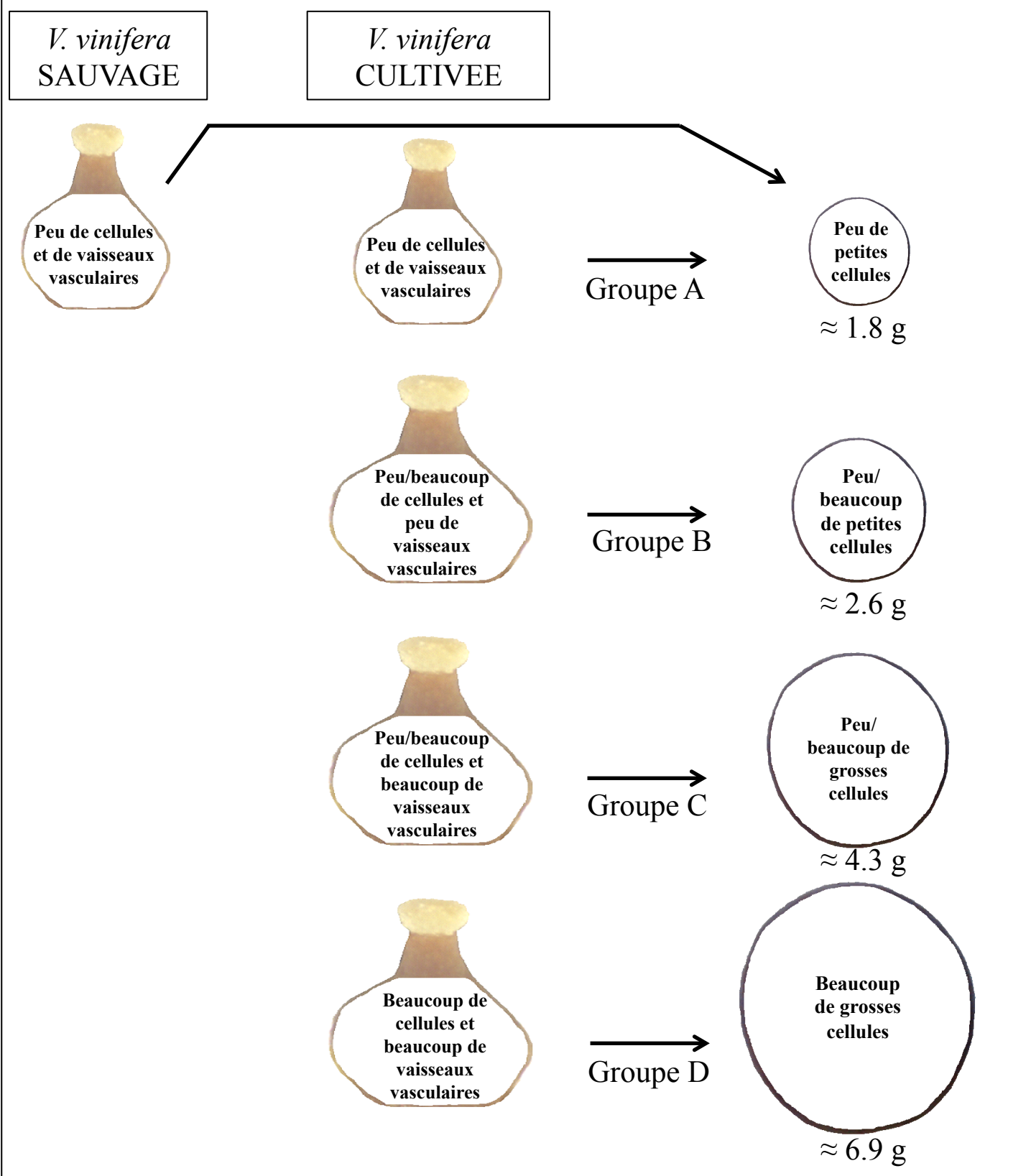
Bien que les génotypes sauvages présentent une diversité nucléotidique plus faible que celle des cultivées, aucun génotype sauvage n'est homozygote pour la région *Flb*, ces résultats suggèrent que cette caractéristique est spécifique des vignes cultivées. Par ailleurs, cette région présente une diversité nucléotidique plus faible chez les génotypes sauvages que le reste du génome, avec de nombreux fragments de gènes séquencés ne présentant aucun polymorphisme et donc une structure de polymorphisme en patchwork. Il est donc possible que des forces évolutives agissent également sur cette région.



## 6. Discussion générale et perspectives



*Photographie de grappes de raisins à maturité.*



**Figure 6.1:** Schéma explicatif des paramètres de l'ovaire et de la baie à l'origine de la variation du poids de la baie chez la vigne *Vitis vinifera* L.

La taille de la baie est un facteur essentiel que ce soit pour la qualité des raisins de table ou celle des raisins de cuve (Champagnol, 1998). Néanmoins, les mécanismes développementaux sous-jacents à sa variation comme le nombre de cellules, la taille des cellules ou le nombre de carpelles n'ont pas été étudiés jusqu'à présent chez la vigne, ou alors seulement chez un nombre très restreint de génotypes ne représentant pas la gamme de variation naturelle de la taille de la baie (Fernandez et al, 2006a et 2006b). De même, les bases génétiques de la variation de ce caractère restent encore largement inconnues et aucun gène impliqué dans ce caractère n'a encore été identifié à ce jour. Or ce type d'information permettrait le développement de méthodes de sélection assistée par marqueurs efficaces pour ce caractère.

Les objectifs de mon travail de thèse ont donc été de contribuer à l'identification des déterminants de la variation de la taille de la baie chez *Vitis vinifera* L. Pour cela, trois stratégies ont été utilisées :

- l'étude morphologique de différents paramètres de l'ovaire et de la baie dans une large collection de génotypes contrastés pour la taille de la baie
- la recherche d'associations statistiques entre les caractères les plus pertinents identifiés précédemment et des gènes candidats pour la variation de la taille de la baie
- une contribution à un effort de cartographie fine du gène *Flb* avec l'étude de la variabilité génétique au locus *Fleshless berry* entre des génotypes sauvages et cultivés afin de rechercher des traces de domestication en relation avec la variation de la taille de la baie.

Les résultats que j'ai obtenus permettent d'avancer sur plusieurs questions que nous nous posions au début de ma thèse, mais ouvrent également beaucoup de questions que je vais détailler maintenant.

### **Quels sont les mécanismes développementaux, génétiques et moléculaires impliqués dans la variation de la taille de la baie chez la vigne ?**

L'étude de différents paramètres de l'ovaire et de la baie, dans un large échantillon de génotypes représentatif de la gamme de variation de la taille de la baie chez la vigne cultivée, a permis de montrer que le poids de la baie était déterminé précocement par le nombre de cellules de l'ovaire avant l'anthèse et par la capacité d'expansion des cellules du mésocarpe à l'origine de la pulpe après l'anthèse (Figure 6.1). Ces résultats sont similaires à ce qui est

connu chez la tomate (Frary et al, 2000; Bertin et al, 2003), l'abricot (Jackson et Coombe, 1966), mais également chez le shorgo (Yang et al, 2009), suggérant un mécanisme très conservé au cours de l'évolution chez les fruits aussi bien chez les dicotylédones que les monocotylédones.

Chez la tomate, un autre caractère joue un rôle essentiel dans la variation de la taille du fruit, à savoir le nombre de carpelles (Tankley, 2004; Cong et al, 2008). Le nombre de carpelles par ovaire varie peu chez la vigne (2 ou 3 carpelles) et cette étude n'a pas permis de relier ce paramètre à la variation de la taille de la baie, bien que la fréquence des génotypes présentant des ovaires à 3 carpelles soit beaucoup plus importante chez les génotypes produisant de grosses baies (65% des génotypes produisent des baies d'environ 4 g et plus). Cependant, il est important de noter qu'aucun des génotypes étudiés ne procédait systématiquement d'ovaire à trois carpelles, mais une fréquence d'ovaire à trois carpelles plus ou moins importante. Le nombre de carpelles par ovaire est donc certainement fortement influencé par des facteurs environnementaux. On trouve les mêmes résultats chez la tomate, puisque chaque variété est définie par un nombre moyen de carpelles (Cheniclet et al, 2005). De plus, dans cette étude je ne me suis intéressée qu'à des génotypes diploïdes. Or certaines variétés anciennes de raisin de table comme *cv. Léopold III* et *cv. Lady Hastings* et des variétés japonaises plus récentes vraisemblablement créées dans les années 1970 (*cv. Takao 4N*, *cv. Kyoho 4N*, *cv. Koshu 4N*) sont tétraploïdes, et ont été créés dans le but d'augmenter la taille de la baie. De façon intéressante, ces variétés montraient également un nombre de carpelles plus important que les variétés diploïdes (T. Lacombe communication personnelle). Il serait donc intéressant d'étudier ce caractère dans un échantillon beaucoup plus important de variétés de table et de déterminer la nature de la corrélation entre ce paramètre, la ploïdie et la variation de la taille de la baie. En cas de démonstration d'une corrélation entre le poids de la baie et le nombre moyen de carpelles dans un échantillon plus important de variétés diploïdes, il serait alors également intéressant de rechercher à nouveau une association entre du polymorphisme de séquence dans l'orthologue putatif de *Fasciated* et la variation du nombre moyen de carpelles dans cette nouvelle collection.

Si les mécanismes développementaux et cellulaires à l'origine de la variation de la taille des fruits sont relativement conservés, il devient intéressant de développer des approches de type « gène candidat » pour commencer à approcher la connaissance des mécanismes moléculaires sous-jacents. Ainsi, nous avons voulu tester si des orthologues putatifs (basés sur une similarité de séquence) de gènes de tomate ou d'*Arabidopsis* impliqués

dans la variation de la taille du fruit ou dans le développement de l'ovaire pouvaient être associés à la variation de la taille de la baie chez la vigne. Nous avons au préalable vérifié, lorsque les données étaient disponibles, si ces gènes étaient bien exprimés dans les ovaires et les baies de vigne.

Nous avons effectivement trouvé une association significative entre deux SNP dans un gène codant pour une putative 3-hydroxy-3-methyl-glutaryl-CoA réductase (*VvHMGR1*), dont un non synonyme dans le site de tétramérisation de l'enzyme, et la variation de la longueur de la baie. Les HMGR sont des enzymes impliquées dans la voie de biosynthèse de l'isoprénolide précurseur de nombreuses phytohormones telles que les cytokinines, l'acide gibbérellique et l'acide abscissique. Chez *Arabidopsis* (Suzuki et al, 2004), l'insertion d'un ADN de transfert dans le gène *HMGR1* provoque un nanisme de la plante du à une inhibition de l'élongation cellulaire, alors que la même mutation dans le second gène de la famille, *HMGR2*, ne semble pas avoir d'effet sur le phénotype. Plusieurs études fonctionnelles mais ne faisant pas intervenir des plantes mutantes pour ces gènes ont montré ou fortement suggéré l'implication de ces enzymes dans la variation de la taille du fruit chez la tomate (Narita and Gruissem, 1989), l'avocat (Cowan et al, 1997) et le melon (gène *Cm-HMGR* orthologue putatif de *HMGR1*; Kobayashi et al, 2002) *via* un contrôle du nombre de cellules du fruit. Chez la vigne, le gène *VvHMGR1*, orthologue putatif de *HMGR1*, appartient à une famille de trois gènes, dont les deux autres ne semblent pas associés à la variation de paramètres de l'ovaire ou de la baie. L'association de *VvHMGR1* avec la longueur des baies et non de leur poids, comme cela semble être le cas chez les autres fruits, peut s'expliquer par le fait que le poids de la pulpe est une variable composite du nombre de cellules et de la taille des cellules, l'absence d'association entre ce gène et la variation du poids de la baie, ne veut pas forcément dire que ce gène n'est pas impliqué dans la variations de ces sous-caractères. De plus chez la tomate, plusieurs QTL sont responsables de la variation du ratio longueur du fruit versus largeur du fruit indépendamment du poids (Tankley, 2004). Les gènes sous-jacents à deux de ces QTL ont été identifiés, *Ovate* crée une élongation asymétrique du fruit, l'élongation du haut du fruit est plus importante que celle du bas et aboutit à un fruit pyriforme (Liu et al, 2002), alors que *Sun* à un effet plus uniforme, aboutissant à un fruit de forme ovoïde (Xiao et al, 2008). Chez la vigne, *VvHMGR1* pourrait être impliqué dans le contrôle de l'élongation cellulaire ou des divisions cellulaires de façon directionnelle, créant ainsi une asymétrie de la forme des baies sans modifier leur poids. L'application de mévastatine, un inhibiteur des HMGR, à un stade précoce du développement de la baie permettrait de mieux comprendre le rôle de cette famille de gènes chez la vigne. Des coupes histologiques de baies ayant ou non reçu de la mévastatine,

permettraient peut-être de trancher sur l'implication de *VvHMGR1*, *via* un contrôle asymétrique de l'élongation cellulaire ou des divisions cellulaires, sur la variation de la longueur de la baie.

Le poids de la baie étant dépendant à la fois du nombre et de la taille des cellules de la pulpe, et la proportion de chaque sous-caractère étant dépendante du génotype, lorsque j'ai recherché des associations entre des SNP dans des gènes candidats et la variation du poids de la pulpe, les gènes impliqués spécifiquement dans l'un de ces deux sous-caractères peuvent ne pas être associés avec la variation du poids de la pulpe dans la collection entière. Ceci pourrait peut être expliquer l'absence d'association entre l'orthologue putatif de gène *Fw2.2* avec la variation du poids de la baie chez la vigne. En effet, le gène *Fw2.2* code pour un régulateur négatif de la division cellulaire qui contribue aux différences de taille du fruit chez la tomate (Frary et al, 2000). Ces études se sont basées sur des descendance s ségrégant uniquement pour le nombre de cellules et non pour le volume cellule ou le nombre de carpelles, dès lors le poids du fruit était corrélé au nombre de cellules (Frary et al, 2000 ; Cong et al, 2002). Alors que dans notre échantillon de vigne, le nombre et la taille des cellules de la baie ségrégent et expliquent son poids final. De plus, la surexpression de l'orthologue putatif de ce gène chez le maïs provoque une diminution de la taille de tous les organes de la plante *via* une diminution du nombre de cellules (Guo et al, 2010), alors que chez le soja, il semble jouer un rôle dans le nombre de nodules *via* une régulation des divisions cellulaires (Libault et al, 2010). Enfin, chez l'avocat, des fruits contrastés pour leur taille en raison d'une différence du nombre de cellules montrent une différence d'expression de ce gène allant dans le sens d'un contrôle négatif de la division cellulaire par ce gène (Dahan et al, 2010). Le rôle de ce gène dans le contrôle de la taille des organes et plus particulièrement du fruit par une régulation négative du nombre de cellules semble donc très conservé aussi bien chez les dicotylédones que les monocotylédones. Le phénotypage du nombre et de la taille des cellules de la baie permettrait certainement de mieux comprendre le rôle de cette famille de gènes chez la vigne et permettrait également d'identifier de nouveaux gènes impliqués dans la variation du poids de la baie.

Pour s'affranchir de l'approche gène candidat basée sur des *a priori*, et ainsi identifier de nouveaux gènes potentiellement impliqués dans la variation de la taille de la baie, une première approche pourrait consister en étude WGAS sur l'échantillon que j'ai utilisé lors de ma thèse, soit en utilisant des SNP sur tout le génome identifiés lors du projet SNPGrpMap (fragments de séquences répartis environ tous les 500 kb dans la génome) et du projet trilatéral GrapeReSeq en cours (re-séquençage Illumina d'une trentaine de *V. vinifera*), soit

dans un futur proche en génotypant directement par séquençage. Une approche complémentaire consisterait à phénotyper pour ces sous-caractères (nombre et taille des cellules de la baie) les populations en ségrégations actuellement disponibles. En effet, des coupes histologiques réalisées lors de ma thèse mais non présentées dans ce manuscrit sur les ovaires de 18 descendants d'une population Syrah x Grenache contrastés pour le poids de la baie (poids de baies moyens variant de 0.9 g à 3.2 g; A. Doligez résultats non publiés), montrent que le nombre de cellules de l'ovaire est variable entre les individus. Une partie des caractères que j'ai mesurés est déjà disponible sur cette population et ce travail pourrait être complété. Il serait également très intéressant de générer une population en ségrégation issue du croisement entre un génotype du Groupe A et un autre du Groupe D (Figure 6.1, les variétés Syrah et Grenache appartenant toutes les deux au Groupe B). Cette population ségrégerait donc pour le nombre et la taille des cellules de la baie. Le choix des parents pourrait également se porter sur des différences de taux de sucre et d'acidité de la baie ou également sur le nombre de vaisseaux conducteurs présents dans l'ovaire (puisque nous avons vu qu'il semblait associé à la variation de la taille de la baie), permettant d'identifier des co-localisations entre des QTL pour la variation du volume cellulaire et des QTL pour le taux de sucre, d'acidité et/ou le nombre de vaisseaux conducteurs. Une étude de génétique d'association pour des gènes candidats co-localisant avec les QTL détectés devrait par la suite aider à l'identification des gènes sous-jacents (Fournier-Level et al, 2009; Emanuelli et al, 2010 ; chapitre 4).

Bien que particulièrement pertinent, le phénotypage du nombre et de la taille des cellules de l'ovaire et de la baie chez la vigne reste pour l'instant très lourd et donc irréalisable sur un échantillon aussi important qu'une descendance F1 ou un panel pour la génétique d'association, et constitue donc un facteur limitant pour l'étude du déterminisme génétique de la variation de la taille de la baie. En effet, je n'ai pas pu mettre au point de méthode haut débit fiable de quantification du nombre de cellules de la baie et je n'ai donc pas pu utiliser ce caractère dans les études de génétique d'association que j'ai menées.

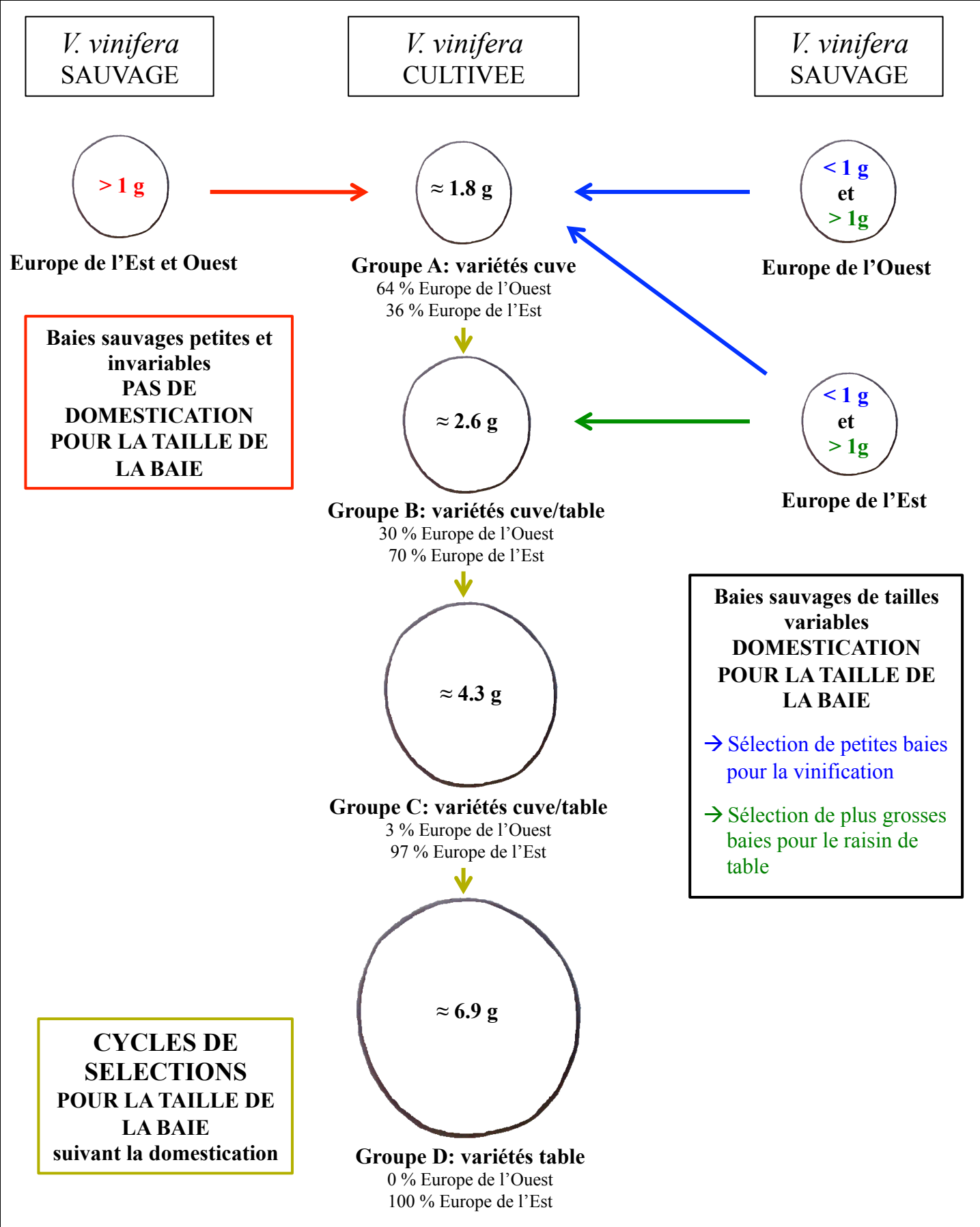
Avec l'arrivée des méthodes de séquençage tout génome à bas coût et des méthodes de génotypage à grande échelle, le phénotypage de grandes séries d'individus constitue donc un verrou majeur dans l'étude des relations génotype-phénotype et pas seulement pour les sous-caractères de la taille de la baie. Le phénotypage est souvent limité par la capacité en temps de travail humain à réaliser les observations dans le temps limité imposé par le stade à caractériser. Cependant, l'augmentation de la capacité à collecter des données sur des





végétaux dans un temps limité tout en améliorant la nature et la précision des mesures est d'un intérêt primordial pour identifier la fonction de nombreux gènes mais également pour tester des gènes et des phénotypes dans des conditions environnementales variées. C'est dans ce sens que de nombreuses équipes mettent actuellement au point des techniques de phénotypage haut-débit notamment grâce aux progrès de la robotique et de l'imagerie (Granier al, 2006; Montes et al, 2007; Golzarian et al, 2011). Chez la vigne, des méthodes et plateformes se mettent également en place en tenant compte des difficultés liées aux plantes pérennes, mais rien concernant le comptage cellulaire de la baie pour le moment.

Enfin, il est important de noter que les études de phénotypage et de génétique d'association n'ont été conduites que sur une année de phénotypage. Or, l'étude d'un grand nombre de génotypes diversifiés pourrait compenser partiellement le manque de répétitions en « tamponant » les effets d'interaction génotype/milieu. De fait, les corrélations inter-annuelles que nous avons obtenu sur nos génotypes contrôles nous ont permis de rassembler les données de phénotypage obtenues sur deux années successives. Cependant, la prise en compte de l'environnement permettrait d'intégrer une composante environnementale dans le modèle d'analyse avec notamment l'intégration d'un effet millésime et d'un effet emplacement au travers du modèle mixte, permettant d'ajuster les relations génotype/phénotype. En effet, le déficit hydrique a une influence majeure sur la taille de la baie avant véraison en diminuant la taille des cellules (Ojeda et al, 2001), mais également l'ensoleillement (Dokoozlian et Kliewer, 1996) et la température qui, lorsqu'elle dépasse 35°C au moment de la floraison, réduit irréversiblement la taille des baies (Kliever, 1977). Or ces facteurs sont hétérogènes selon l'orientation des vignobles et leurs effets doivent être estimés dans des essais randomisés. Cela peut s'avérer difficile à réaliser en termes de faisabilité chez les pérennes en particulier. En attendant, une seconde année de phénotypage des 3 sous-populations sur le Domaine expérimental de Vassal (Marseillan, France) pour les caractères de l'ovaire et de la baie permettrait de renforcer toutes les données obtenues pendant ma thèse. Par la suite, cette collection ayant été surgreffée à un autre emplacement (Domaine expérimental du Chapitre, Villeneuve les Maguelones, France), le phénotypage de cette collection sur les deux sites permettrait une meilleure estimation de la composante environnementale sur chaque caractère phénotypique.



**Figure 6.2:** Schéma hypothétique de la domestication de la vigne *Vitis vinifera* L. ayant porté sur la taille des baies en fonction de leur utilisation et de leur origine géographique (Europe de l'Est et Ouest), et cycles de sélection pour la taille de la baie ayant suivi la domestication.

### **Est-ce que la taille de la baie a été un des caractères sélectionnés pendant la domestication de la vigne ?**

La domestication d'une espèce s'accompagne de modifications morphologiques majeures des organes d'intérêts pour l'Homme *via* la sélection indirecte de gènes impliqués dans ces caractères. Par exemple, l'ancêtre sauvage putatif de la tomate cultivée porte des fruits ne pesant que quelques grammes. En revanche, un seul fruit d'une variété de tomate moderne peut peser jusqu'à 1 kg, ce qui représente une augmentation du poids de près de 1 000 fois (Frary et al, 2000; Tankley, 2004; Cong et al, 2008). Un des gènes responsables de cette différence et impliqué dans la regulation du nombre de cellules est *Fw2.2* (Frary et al, 2000). Une situation similaire existe pour la plupart des autres espèces fruitières domestiquées (Smartt et Simmonds, 1995). D'après Levadoux (1956), et mon étude une différence de taille de baie est également observée chez la vigne entre le compartiment sauvage (baie <1 g) et cultivée (baie pesant jusqu'à 10 g et plus). Cependant, l'étude morphologique des paramètres de l'ovaire et de la baie chez le génotype sauvage que j'ai phénotypé a soulevé l'hypothèse que la taille de la baie pourrait ne pas être un des caractères sélectionnés par l'Homme lors de la domestication de la vigne. En effet, les ovaires et baies du génotype sauvage que nous avons étudiés ont les mêmes caractéristiques (nombre et taille des cellules de l'ovaire et de la baie) que celles de certains génotypes cultivés à petites baies. Deux hypothèses peuvent expliquer ces résultats (Figure 6.2):

- soit tous les *V. vinifera* sauvages produisaient de petites baies en raison d'une forte pression de sélection (naturelle) sur la taille de la baie. En effet, les petites baies permettent une meilleure dispersion des grains car elles peuvent facilement être cueillies, transportées et dispersées par les rongeurs et les oiseaux. Dans ce cas l'Homme n'avait pas le choix et la sélection a uniquement porté sur d'autres critères comme les fleurs hermaphrodites, le nombre de baies par grappe, le nombre de grappes par pied, l'augmentation des sucres... (This et al, 2006). L'augmentation de la taille de la baie aurait été un critère de sélection ultérieur, après domestication, en relation avec une diversification de l'utilisation des variétés et l'apparition de mutations. Ceci serait compatible avec l'observation que certains cépages très proches morphologiquement des vignes sauvages comme le Riesling, le Pinot et le Petit Verdot soient également caractérisés par une valeur vinique élevée et souvent par une saveur typique des baies qui rappelle celle des baies des lambrusques (Levadoux, 1956). Par ailleurs, le fait que la taille de la baie n'ait pas été un caractère sélectionné pendant la domestication de la vigne, pourrait également expliquer que les traces de

sélection identifiées dans la région du gène *Flb* entre des vignes sauvages et cultivées ne soient pas associées à la variation de la taille de la baie.

- soit la taille de la baie était bien variable au sein du compartiment sauvage, et il y aurait pu y avoir deux types de domestication. En effet, la domestication de la vigne est étroitement liée à l'apparition de la vinification (McGovern, 2003). l'Homme aurait donc sélectionné des génotypes à petites baies produisant des vins plus gustatifs et de meilleure qualité que ceux réalisés à partir de plus grosses baies (Cordonnier, 1976). Les variétés cultivées de notre étude présentant de très petites baies (<1g) similaires à celle des vignes sauvages sont toutes originaires de l'Ouest de l'Europe (Figure 6.2) où la vigne était cultivée très majoritairement pour la vinification. De plus, la variation de la taille des baies des variétés de cuve est plus importante chez les variétés originaires de l'Est de l'Europe que chez celles originaires de l'Ouest. De même, l'ensemble des variétés de table, produisant les plus grosses baies, sont originaires de l'Est de l'Europe, suggérant une évolution différentielle dans ces deux régions. Par ailleurs, les études visant à décrire la structure génétique de la vigne cultivée soulignent toutes que le principal facteur de différenciation chez la vigne cultivée est l'origine géographique Est/Ouest des cépages, suggérant également une sélection divergente des cultivars dans les deux régions (Arroyo-Garcia et al, 2006; This et al, 2006; Myles et al, 2011). Il est donc tout à fait possible qu'en Europe de l'Est et au Moyen-Orient l'Homme ai sélectionné des génotypes avec différentes tailles de la baie selon leur utilisation dès la domestication. Des cycles de sélection ultérieurs ont permis d'obtenir les génotypes extrêmes actuels.

Dans notre étude nous n'avons pu étudier qu'un seul génotype sauvage (Lambrusque Grésigny I), les autres génotypes sauvages à ma disposition étant soit des mâles, soit trop jeunes pour produire des baies, soit avaient subi l'attaque de rongeurs. Ce génotype étant originaire de France donc de l'Ouest de l'Europe, il serait indispensable de répéter cette étude sur d'autres sauvages et plus particulièrement sur des génotypes originaires de l'Est de l'Europe afin d'avoir une connaissance de la variation des caractères que nous avons mesurés dans le compartiment sauvage actuel, ce qui permettrait d'éclairer nos hypothèses. La détection de QTL de poids de la baie dans une population issue du croisement entre un génotype sauvage et cultivé permettrait également d'éclairer la contribution possible du compartiment sauvage à la variation de ce caractère (Doebley et al, 1995 ; Frary et al, 2000). Une autre possibilité serait de mener en parallèle des études WGAS pour la taille de la baie à

la fois dans le compartiment sauvage, si ce caractère s'avère variable et dans le compartiment cultivé mais ceci nécessiterait de mieux connaître la diversité génétique du compartiment sauvage afin de constituer des panels d'étude adéquats.

Une partie de ma thèse a porté sur l'identification de traces de domestication dans la région *Fleshless berry* supposée être impliquée dans la variation de la taille de la baie, à travers l'analyse de la diversité génétique entre un set de génotypes sauvages et un de cultivés. Cette étude a révélé comme d'autres précédemment (Aradhya et al, 2003; Arroyo-Garcia et al, 2006; Grassi et al, 2008 ; Myles et al, 2011 ; Laucou et al, 2011), une diversité génétique du compartiment sauvage plus faible que celle du compartiment cultivé, permettant seulement l'identification de traces de sélection purificatrice, dans le cadre de panels de génotypes sauvages non garantis pour être représentatifs de la diversité totale du compartiment. En effet, les gènes sous sélection purificatrice sont caractérisés par une diversité génétique plus importante chez la vigne sauvage que chez la vigne cultivée, alors que les gènes soumis à la sélection diversificatrice sont caractérisés par une diversité génétique plus importante chez la vigne cultivée que chez la vigne sauvage. Par ailleurs, on sait que la vigne cultivée a subi une re-diversification après sa domestication à l'origine de sa grande diversité génétique actuelle. Ceci a été illustré récemment lors de l'étude de l'évolution des gènes *VvMYBA* impliqués dans la voie de biosynthèse des anthocyanes qui présente une très forte diversification dans le compartiment cultivé (Fournier-Level et al, 2010). En utilisant les ressources de génotypes sauvages actuelles, on risque donc de passer à côté de nombreux gènes impliqués dans des caractères d'intérêts agronomiques.

La diminution de la diversité génétique du compartiment sauvage pourrait être liée à l'introduction des agents pathogènes en Europe (*Plasmopara viticola*, *Erysiphe necator* et *Phylloxera vastatrix*) qui ont détruit une grande partie des populations de vignes sauvages européennes à la fin du XIXème siècle. La diversité de *V. vinifera* ssp *sylvestris* n'est très certainement aujourd'hui que le pâle reflet de ce qui existait avant l'introduction de ces agents pathogènes (Arnold et al, 1998 voir Roberto, Pouget, 1990), lui valant d'être placée dans la liste rouge des espèces en voie d'extinction par l'Union Internationale pour la Conservation de la Nature (UICN). De plus, les faibles connaissances actuelles du compartiment sauvage et la confusion possible avec certains génotypes cultivés (un retour à l'état sauvage de certaines vignes cultivées redonne des vignes identiques au type *silvestris*; Levadoux, 1956) rendent la recherche de traces de domestication/sélection périlleuse chez la vigne.

D'ailleurs, bien qu'une recherche sur le génome entier de traces de sélection ait été réalisée

très récemment chez la vigne (Myles et al, 2011) entre des génotypes sauvages et cultivés, cette étude n'a permis d'identifier qu'une seule trace de sélection. Cette trace correspondait à une diminution de la diversité nucléotidique chez les génotypes cultivés par rapport aux sauvages sur 5 Mb au niveau du chromosome 17, et n'a pour le moment pas été reliée à une variation phénotypique distinguant ces deux sous-espèces. Or, on sait que le principal caractère ciblé pendant la domestication de la vigne est le passage de fleurs dioïques à hermaphrodites (This et al, 2006), le QTL responsable a été cartographié dans plusieurs études indépendantes sur le chromosome 2 (Dalbo et al, 2000; Lowe et Walker, 2006 ; Riaz et al, 2006 ; Marguerite et al, 2009 ; Chaïb et al, 2010). Myles et al (2011) auraient théoriquement dû trouver une trace de sélection sur ce chromosome. Ces résultats peuvent être liés à plusieurs facteurs : le mauvais échantillonnage du compartiment sauvage qui est encore mal connu ou le faible DL chez la vigne associé à un trop faible nombre de marqueurs génotypés dans cette étude.

Il est donc très important dans un futur proche de mieux caractériser le compartiment sauvage, afin de pouvoir identifier les gènes de la domestication. Il se peut que chez la vigne le compartiment sauvage actuel en raison de sa petite taille et de son évolution, ne permette pas de rechercher de traces de domestication.

## 7. Références bibliographiques



*Source: <http://www.vignoblesdusud.com>*





**Les références bibliographiques regroupent l'ensemble des références du manuscrit y compris celles associées aux articles.**

**Abdallah JM, Goffinet B, Cierco-Ayrolles C, Pérez-Enciso M.** Linkage disequilibrium fine mapping of quantitative trait loci: a simulation study. *Genet Sel Evol* 2003, 35:513-32.

**Adam-Blondon A-F, Jaillon O, Vezzulli S, Zharkikh A, Troggio M, Velasco R.** Genetics, genomics and breeding of grapes. Chapter 9 : Genome Sequence Initiatives. In: Science Publishers, P.O. Box 699, Enfield, NH 03748, USA, 2011, pp:211-234 in press.

**Adam-Blondon AF, Roux C, Claux D, Butterlin G, Merdinoglu D, This P.** Mapping 245 SSR markers on the *Vitis vinifera* genome: a tool for grape genetics. *Theor Appl Genet* 2004, 109:1017-1027.

**Alleweldt G, Dettweiler E.** The genetic resources of *Vitis*: world list of grapevine collections. 1994.

**Alleweldt G, Koch R.** Ethylene content in ripening grape berries. *Vitis* 1977, 16:263-271.

**Altshuler D, Daly MJ, Lander ES.** Genetic mapping in human disease. *Science* 2008, 322:881-888.

**Andersen PC, Brodbeck BV, Mizell RF.** Diurnal variations in tension, osmolarity and the composition of nitrogen and carbon assimilates in xylem fluid of *Prunus persica*, *Vitis* hybrid, and *Pyrus communis*. *J Amer Soc Hort Sci* 1995, 120:600-606.

**Angenent GC, Colombo L.** Molecular control of ovule development. *Trends Plant Sci* 1996, 1: 228-232.

**Angenent GC, Franken J, Busscher M, Van Dijken A, Van Went JL, Dons HJM, Van Tunen AJ.** A novel class of MADS box genes is involved in ovule development in petunia. *Plant Cell* 1995, 17:569-1582.

**Aradhya MK, Dangl GS, Prins BH, Boursiquot J-M, Walker MA, Meredith CP, Simon CJ.** Genetic structure and differentiation in cultivated grape, *Vitis vinifera* L. *Genet Res Camb* 2003, 81:179-192.

**Arnold C, Gillet F, Gobat JM.** Occurrence of the wild vine *Vitis vinifera* ssp. *silvestris* in Europe. *Vitis* 1998, 37:159-170.

**Arrigo N, Arnold C.** Naturalised *Vitis* rootstocks in Europe and consequences to native wild grapevine. *PLoS ONE* 2007, 2:e521.

**Arrigo N, Arnold C.** Naturalised *Vitis* rootstocks in Europe and consequences to native wild grapevine. *PLoS ONE* 2007, 2:e521.

**Arroyo-Garcia R, Ruiz-Garcia L, Bolling L, Ocete R, Lopez MA, Arnold C, Ergul A, Söylemezoglu G, Uzun HJ, Cabello F, Ibanez J, Aradhya MK, Atanassov I, Balint S, Cenis JL, Costantini L, Gorislavets S, Grando MS, Klein BY, Govern PEM, Merdinoglu D, Pejic I, Pelsy F, Primikiris N, Risovannaya V, Roubelakis-Angelakis KA, Snoussi H,**

**Sotiri P, Tamhankar S, This P, Troshin L, Malpica JM, Lefort F, Martinez-Zapater JM.** Multiple origins of cultivated grapevine (*Vitis vinifera* L. ssp. *sativa*) based on chloroplast DNA polymorphisms. *Molecular Ecology* 2006, 15:3707-3714.

**Atwell S, Huang YS, Vilhjálmsson BJ, Willems G, Horton M, Li Y, Meng D, Platt A, Tarone AM, Hu TT, Jiang R, Mulyati NW, Zhang X, Amer MA, Baxter I, Brachi B, Chory J, Dean C, Debieu M, de Meaux J, Ecker JR, Faure N, Kniskern JM, Jones JD, Michael T, Nemri A, Roux F, Salt DE, Tang C, Todesco M, Traw MB, Weigel D, Marjoram P, Borevitz JO, Bergelson J, Nordborg M.** Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* 2010, 465:627-31.

**Baldet P, Hernould M, Laporte F, Mounet F, Just D, Mouras A, Chevalier C, Rothan C.** The expression of cell proliferation-related genes in early developing flowers is affected by fruit load reduction in tomato plants. *J Exp Bot* 2006, 57:961-970.

**Banerjee MK and Kalloo G.** The inheritance of earliness and fruit weight in crosses between cultivated tomatoes and two species of *Lycopersicon*. *Plant Breed* 1989, 102:148-152.

**Barnaud A, Lacombe T and Doligez A.** Linkage disequilibrium in cultivated grapevine, *Vitis vinifera* L. *Theor Appl Genet* 2009, 112:708-716.

**Barnaud A, Lacombe T, Doligez A.** Linkage disequilibrium in cultivated grapevine, *Vitis vinifera* L. *Theor Appl Genet* 2006, 112:708-16.

**Barnavon L, Doco T, Terrier N, Ageorges A, Romieu C, Pellerin P.** Involvement of pectin methyl-esterase during the ripening of grape berries: partial cDNA isolation, transcript expression and changes in the degree of methyl-esterification of cell wall pectins. *Phytochemistry* 2001, 58:693-701.

**Battilana J, Costantini L, Emanuelli F, Sevini F, Segala C, Moser S, Velasco R, Versini G, Grando MS.** The 1-deoxy-D-xylulose5-phosphate synthase gene co-localizes with a major QTL affecting monoterpene content in grapevine. *Theor Appl Genet* 2009, 118:653-69.

**Benharbit El Alami N.** Recherches sur le développement normal et anormal du gynécée de *Vitis vinifera* L. du bouton floral à la fleur fécondée. Importance des tanins vacuolaires et relation avec la coulure. Thèse de l'Université de Bordeaux II, France, 1995, 126p.

**Bergervoet JHW, Verhoeven HA, Luun JW, Gilissen JW, Bino RJ.** High amounts of nuclear DNA in tomato (*Lycopersicon esculentum* Mill.) pericarp. *Plant Sci* 1996, 116:141-145.

**Bertin N, Borel C, Brunel B, Cheniclet C, Causse M.** Do genetic makeup and growth manipulation affect tomato fruit size by cell number, or cell size and DNA endoreduplication? *Annals of Botany* 2003, 92:415-424.

**Bertin N, Causse M, Brunel B, Tricon D, Genard M.** Identification of growth processes involved in QTL for tomato fruit size and composition. *J Exp Bot* 2009, 60:237-248.

**Bessis R, Fournioux JC.** Zone d'abscission et coulure de la vigne. *Vitis* 1992, 31: 9-21.

**Bohner J, Bangerth F.** Effects of fruit-set sequence and defoliation on cell number, cell-size and hormone levels of tomato fruits (*Lycopersicon esculentum* mill) within a truss. *Plant growth regulation* 1988, 7:141-155.

**Boss PK, Davies C, Robinson SP** Anthocyanin composition and anthocyanin pathway gene expression in grapevine sports differing in berry skin colour. *Aust J Grape Wine Res* 1996, 2: 163-170.

**Boss PK, Sensi E, Hua C, Davies C, Thomas MR.** Cloning and characterization of grapevine (*Vitis vinifera* L.) MADS-box genes expressed during inflorescence and berry development. *Plant Sci* 2002, 162:887-895.

**Boss PK, Thomas MR.** Association of dwarfism and floral induction with a grape 'green revolution' mutation. *Nature* 2002, 416: 847-850.

**Boss PK, Vivier M, Matsumoto S, Dry IB, Thomas MR.** A cDNA from grapevine (*Vitis vinifera* L.), which shows homology to AGAMOUS and SHATTERPROOF, is not only expressed in flowers but also throughout berry development. *Plant Mol Biol* 2001, 45:541-553.

**Botta R, Vellania R, Me G.** Grapevine breeding by gamma radiations: results and prospects. *Acta Hort* 1987, 224: 30-36.

**Boubals D, Bourzeix M, Guitraud J.** Gora Chirine, variete de vigne iranienne à faible teneur en acides organiques dans les baies. *Ann Amelior Plantes* 1971, 21: 281-285.

**Bouquet A.** Genetics, genomics and breeding of grapes. Chapter 1 : Grapevines and Viticulture. In: Science Publishers, P.O. Box 699, Enfield, NH 03748, USA, 2011, pp:1-29 in press.

**Bouquet A.** Origine et évolution de l'encépagement français à travers les siècles. *Progrès Agricole et Viticole* 1982, 5:110-121.

**Boursiquot JM, Dessup M, Rennes C.** Distribution des principaux caractères phénologiques, agronomiques et technologiques chez *Vitis vinifera* L. *Vitis* 1995, 34:31-35.

**Boursiquot J-M, Lacombe T, Laucou V, Julliard S, Perrin F-X, Lanier N, Legrand D, Meredith CP, This P.** Parentage of Merlot and related winegrape cultivars of southwestern France: discovery of the missing link. *Aust J Grape Wine Res* 2009, 15: 144-15.

**Boursiquot JM, This P.** Les Nouvelles techniques utilisées en Ampélographie: informatique et marquage. *Journal International de la Vigne et du Vin* 1996, Hors Série:13- 23.

**Bowers JE, Boursiquot J-M, This P, Chu K, Johansson H, Meredith CP.** Historical genetics: The parentage of Chardonnay, Gamay and other wine grapes of northeastern France. *Science* 1999, 285:1562-1565.

**Bowers JE, Dangl GS, Vignani R, et al.** Isolation and characterization of new polymorphic simple sequence repeat loci in grape (*Vitis vinifera* L). *Genome* 1996, 39:628-633.

**Bowers JE, Meredith CP.** The parentage of a classic wine grape, Cabernet Sauvignon. *Nat Genet* 1997, 16:84-7.

**Bowers JE, Siret R, Meredith CP, This P, Boursiquot J-M.** A single pair of parents proposed for a group of grapevine varieties in northeastern France. *Acta Hort* 2000, 528:129-132.

**Bowman JL, Smyth DR, Meyerowitz EM.** Genetic interactions among floral homeotic genes of Arabidopsis. *Development* 1991, 112:1-20.

**Buckler ES, Holland JB, Bradbury PJ, Acharya CB, Brown PJ, Browne C, Ersoz E, Flint-Garcia S, Garcia A, Glaubitz JC, Goodman MM, Harjes C, Guill K, Kroon DE, Larsson S, Lepak NK, Li H, Mitchell SE, Pressoir G, Peiffer JA, Rosas MO, Rocheford TR, Romay MC, Romero S, Salvo S, Sanchez Villeda H, da Silva HS, Sun Q, Tian F, Upadyayula N, Ware D, Yates H, Yu J, Zhang Z, Kresovich S, McMullen MD.** The genetic architecture of maize flowering time. *Science* 2009, 325:714-8.

**Burke JM, Knapp SJ, Rieseberg LH.** Genetic consequences of selection during the evolution of cultivated sunflower. *Genetics* 2005, 171:1933-1940.

**Cabezas JA, Cervera MT, Ruiz-Garcia L, Carreno J, Martinez-Zapater JM.** A genetic analysis of seed and berry weight in grapevine. *Genome* 2006, 49:1572-1585.

**Calonje M, Cubas P, Martinez-Zapater J-M, Carmona MJ.** Floral meristem identity genes are expressed during tendril development in grapevine. *Plant Physiol* 2004, 135:1491-1501.

**Camus-Kulandaivelu L, Chevin LM, Tollon-Cordet C, Charcosset A, Manicacci D, Tenaillon MI.** Patterns of molecular evolution associated with two selective sweeps in the Tb1-Dwarf8 region in maize. *Genetics* 2008, 180:1107-21.

**Camus-Kulandaivelu L, Veyrieras JB, Madur D, Combes V, Fourmann M, Barraud S, Dubreuil P, Gouesnard B, Manicacci D, Charcosset A.** Maize adaptation to temperate climate: Relationship between population structure and polymorphism in the Dwarf8 Gene. *Genetics* 2006, 172: 2449-2463.

**Carbonneau A, Deloir A, Jaillard B.** La vigne: Physiologie, terroir, culture. Ed Dunod, Pris, France, 2007, 442p.

**Carmona MJ, Chaïb J, Martínez-Zapater JM, Thomas MR.** A molecular genetic perspective of reproductive development in grapevine. *J Exp Bot* 2008, 59:2579-96.

**Carolus M.** Description des stades du développement des primordia inflorescentiels durant l'organogenèse des bourgeons latents de la vigne (*Vitis vinifera* L. cv Merlot). *Conn Vigne et Vin* 1971, 2: 163-173.

**Casa AM, Mitchell SE, Hamblin MT, Sun H, Bowers JE, Paterson AH, Aquadro CF, Kresovich S.** Diversity and selection in sorghum: Simultaneous analyses using simple sequence repeats. *Theor Appl Genetics* 2005, 111: 23-30.

**Causse M, Duffe P, Gomez M-C, Buret M, Damidaux R, Zamir D, Gur A, Chevalier C, Lemaire-Chamley M, Rothan C.** A genetic map of candidate genes and QTL involved in tomato fruit size and composition. *J Exp Bot* 2004, 55:1671-1685.

**Chaïb J, Lecomte L, Buret M, Causse M.** Stability over genetic backgrounds, generations and years of quantitative trait locus (QTL) for organoleptic quality in tomato. *Theor Appl Genet* 2006, 112:934-944.

**Chaïb J, Torregrosa L, Mackenzie D, Corena1 P, Bouquet A, Thomas MR.** The grape microvine – a model system for rapid forward and reverse genetics of grapevines. *The Plant J* 2010, 62:1083-1092.

**Champagnol F.** Elements de physiologie de la vigne et viticulture générale. Ed Dehan, Montpellier, France, 1984, 352p.

**Champagnol F.** Critères de qualité de la vendange. In Oenologie, Fondements Scientifiques et Technologiques. Ed Flanzky C. Lavoisier Tec &Do, Paris, France, 1998, 659p.

**Chapman MA, Pashley CH, Wenzler J, Hvala J, Tang S, Knapp SJ, Burke JM.** A genomic scan for selection reveals candidates for genes involved in the evolution of cultivated sunflower (*Helianthus annuus*). *Plant Cell* 2008, 20:2931-45.

**Chappell J.** Biochemistry and molecular biology of the isoprenoid biosynthetic pathway in plants. *Annu Rev Plant Physiol and Plant Mol Biol* 1995, 46:521-547.

**Chatelet P, Laucou V, Fernandez L, Sreekantan L, Lacombe T, Martinez-Zapater JM, Thomas MR, Torregrosa L.** Characterization of *Vitis vinifera* L. somatic variants exhibiting abnormal flower development patterns. *J Exp Bot*, 2007, 58:4107-18.

**Cheniclet C, Rong WY, Causse M, Frangne N, Bolling L, Carde J-P, Renaudin J-P.** Cell expansion and endoreduplication show a large genetic variability in pericarp and contribute strongly to tomato fruit growth. *Plant Physiol* 2005, 139:1984-1994.

**Chervin C, El-Kereamy A, Roustan J-P., Latché A, Lamon J, Bouzayen M.** Ethylene seems required for the berry development and ripening in grape, a non-climacteric fruit. *Plant Sci* 2004, 167:1301-1305.

**Chevalier C, Nafati M, Mathieu-Rivet E, Bourdon M, Frangne N, Cheniclet C, Renaudin JP, Gévaudant F, Hernould M.** Elucidating the functional role of endoreduplication in tomato fruit development. *Ann Bot* 2011, DOI:10.1093/aob/mcq257.

**Ching A, Caldwell KS, Jung M, Dolan M, Smith OS, Tingey S, Morgante M, Rafalski AJ.** SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. *BMC Genet* 2002, 3:19.

**Cipriani G, Spadotto A, Jurman I, Di Gaspero G, Crespan M, Meneghetti S, Frare N, Vignani R, Cresti M, Morgante M, Pezzotti M, Pe E, Policriti A, Testolin R.** The SSR-based molecular proWle of 1005 grapevine (*Vitis vinifera* L.) accessions uncovers new synonymy and parentages, and reveals a large admixture amongst varieties of diVerent geographic origin *Theor Appl Genet* 2010, 121:1569-1585.

**Clark RM, Linton E, Messing J, and Doebley JF.** Pattern of diversity in the genomic region near the maize domestication gene *tb1*. *PNAS* 2004, 101:700-707.

**Clark RM, Wagler TN, Quijada P, Doebley J.** A distant upstream enhancer at the maize domestication gene *tb1* has pleiotropic effects on plant and inflorescent architecture. *Nat Genet* 2006, 38:594-7.

**Coen ES, Meyerowitz EM.** The war of the whorls: genetic interactions controlling flower development. *Nature* 1991, 353: 31-37.

**Coleman C, Copetti D, Cipriani G, Hoffmann S, Kozma P, Kovacs L, Morgante M, Testolin R, Di Gaspero G.** The powdery mildew resistant gene *REN1* co-segregates with an NBS-LRR gene cluster in two central Asian grapevines. *BMC Genet* 2009, 10:89.

**Cong B, Barrero LS, Tanksley SD.** Regulatory change in YABBY-like transcription factor led to evolution of extreme fruit size during tomato domestication. *Nat Genet* 2008, 40:800-804.

**Cong B, Lui J, Tanksley SD.** Natural alleles at a tomato fruit size quantitative trait locus differ by heterochronic regulatory mutations. *Proc Acad Sci USA* 2002, 99:13606-13611.

**Coombe BG.** Relationship of growth and development to changes in sugars, auxins, and gibberellins in fruit of seeded and seedless varieties of *Vitis vinifera*. *Plant Physiol* 1960, 35:241-250.

**Coombe BG.** Development of the grape berry. I; Effects of time of flowering and competition. *Aust J Agric Res* 1980, 31:125-131.

**Coombe BG.** Distribution of solutes within the developing grape berry in relation to its morphology. *Am J Enol Vitic* 1987, 38:120-127.

**Coombe BG.** Research on the development and ripening of the grape berry. *Am J Enol Vitic* 1992, 43:101-110.

**Coombe BG.** The development of fleshy fruits. *Ann Rev Plant Physiol* 1976, 27:507-528.

**Coombe BG.** The regulation of set and development of the grape berry. *Acta Hort* 1973, 34:261-271.

**Cordonnier E.** Qualité de la vendange et méthodologie de la sélection viticole. *Progr Agric Vitic* 1976, 93:760-762.

**Costantini L, Battilana J, Lamaj F, Fanizza G, Grando M.** Berry and phenology-related traits in grapevine (*Vitis vinifera* L.) From Quantitative Trait Loci to underlying genes. *BMC Plant Biology* 2008, 17:8-38.

**Cowan AK, Moore-Gordon CS, Bertling I, Wolstenholme BN.** Metabolic Control of Avocado Fruit Growth (Isoprenoid Growth Regulators and the Reaction Catalyzed by 3-Hydroxy-3-Methylglutaryl Coenzyme A Reductase). *Plant Physiol* 1997, 114:511-518.

**Dahan Y, Rosenfeld R, Zadiranov V, Irihimovitch V.** A proposed conserved role for an avocado FW2.2-like gene as a negative regulator of fruit cell division. *Planta* 2010, 232:663-76.

**Dalbó MA, Ye GN, Weeden NF, Steinkellner H, Sefc KM, Reisch BI.** A gene controlling sex in grapevine placed on a molecular marker-based genetic map. *Genome* 2000, 43: 333-340.

**Di Gaspero G, Cipriani G, Adam-Blondon AF, Testolin R.** Linkage maps of grapevine displaying the chromosomal locations of 420 microsatellite markers and 82 markers for R-gene candidates. *Theor Appl Genet* 2007, 114:1249-1263.

**Di Vecchi-Staraz M, Bandinelli R, Boselli M, This P, Boursiquot JM, Laucou V, Lacombe T, Vares D.** Genetic structuring and parentage analysis for evolutionary studies in grapevine: Kin group and origin of the cultivar Sangiovese revealed. *Journal of the American Society for Horticultural Science* 2007, 132:514-524.

**Di Vecchi-Staraz M, Laucou V, Bruno G, Lacombe T, Gerber S, Bourse T, Boselli M, This P.** Low level of pollen-mediated gene flow from cultivated to wild grapevine: consequences for the evolution of the endangered subspecies *Vitis vinifera* L. subsp. *silvestris*. *J Heredity* 2009, 100:66-75.

**Díaz-Riquelme J, Lijavetzky D, Martínez-Zapater JM, Carmona MJ.** Genome-wide analysis of MIKCC-type MADS box genes in grapevine. *Plant Physiol* 2009, 149:354-69.

**Ditta G, Pinyopich A, Robles P, Pelaz S, Yanofsky MF.** The SEP4 gene of *Arabidopsis thaliana* functions in floral organ and meristem identity. *Curr Biol* 2004, 14:1935-40.

**Doebley J, Stec A, Gustus C.** Teosinte branched1 and the origin of maize: evidence for epistasis and the evolution of dominance. *Genetics* 1995, 141:333-46.

**Doebley J, Stec A, Hubbard L.** The evolution of apical dominance in maize. *Nature* 1997, 386:485-8.

**Dokoozlian NK, Kliewer WM.** Influence of light on grape berry growth and composition varies during fruit development. *J Amer Soc Hort Sci* 1996, 121:869-874.

**Doligez A, Adam-Blondon AF, Cipriani G, Di Gaspero G, Laucou V, Merdinoglu D, Meredith CP, Riaz S, Roux C, This P.** An integrated SSR map of grapevine based on five mapping populations. *Theor Appl Genet* 2006, 113: 369-382.

**Doligez A, Bouquet A, Danglot Y, Lahogue F, Riaz S, Meredith CP, Edwards KL, This P.** Genetic mapping of grapevine (*Vitis vinifera* L.) applied to the detection of QTL for seedlessness and berry weight. *Theor Appl Genet* 2002, 105:780-795.

**Donald TM, Pellerone F, Adam-Blondon AF, Bouquet A, Thomas MR, Dry IB.** Identification of resistance gene analogs linked to a powdery mildew resistance locus in grapevine. *Theor Appl Genet* 2002, 104: 610–618.

**Donèche B, Chardonnet C.** Evolution et localisation des principaux cations au cours du développement du raisin. *Vitis* 1992, 31:175-181.

**Duchêne E, Butterlin G, Claudel P, Dumas V, Jaegli N, Merdinoglu D.** A grapevine (*Vitis vinifera* L.) deoxy-D-xylulose synthase gene colocates with a major quantitative trait loci for terpenol content. *Theor Appl Genet* 2009, 118:541-52.

**East EM.** A Mendelian interpretation of variation that is napparently continuous. *The American Naturalist* 1910, 44: 65-82.

**Ebadi A, May P, Sedgley M, Coombe BG.** Effect of low temperature near flowering time on ovule development and pollen tube growth in the grapevine (*Vitis vinifera* L.), cvs Chardonnay and Shiraz. *Australian Jorنال of Grape and Wine Research* 1995, 1:11-18.

**Ellstrand NC.** Current knowledge of gene flow in plants: implications for transgene flow. *Philos Trans R Soc Lond B Biol Sci* 2003, 358:1163-70.

**Emanuelli F, Battilana J, Costantini L, Le Cunff L, Boursiquot JM, This P, Grando MS.** A candidate gene association study on muscat flavor in grapevine (*Vitis vinifera* L.). *BMC Plant Biol* 2010, 9:10:241.

**Enjuto M, Balcells L, Campos N, Caelles C, Arro M, Boronat A.** Arabidopsis thaliana contains two differentially expressed 3-hydroxy-3-methylglutaryl-CoA reductase genes, which encode microsomal forms of the enzyme. *Proc Natl Acad Sci USA* 1994, 91:927-931.

**Esau K.** Plant anatomy. New York: Wiley, pp767  
**Fanizza G, Lamaj F, Costantini L, Chaabane R, Grando MS.** QTL analysis for fruit yield components in table grapes (*Vitis vinifera*). *Theor Appl Genet* 1965, 111:658-664.

**Fanizza G, Lamaj F, Costantini L, Chaabane R, Grando MS.** QTL analysis for fruit yield components in table grapes (*Vitis vinifera*). *Theor Appl Genet* 2005, 111:658-664.

**Favaro R, Pinyopich A, Battaglia R, Kooiker M, Borghi L, Ditta G, Yanofsky MF, Kater MM, Colombo L.** MADS-box protein complexes control carpel and ovule development in Arabidopsis. *Plant Cell* 2003, 15: 2603-11.

**Fernandez L, Doligez A, Lopez G, Thomas MR, Bouquet A, Torregrosa L.** Somatic chimerism, genetic inheritance, and mapping of the fleshless berry (flb) mutation in grapevine (*Vitis vinifera* L.). *Genome* 2006c, 49:721-8.

**Fernandez L, Pradal M, Lopez G, Berud F, Romieu C, Torregrosa L.** Berry size variability in *Vitis vinifera* L. *Vitis* 2006a, 45:53-55.

**Fernandez L, Romieu C, Moing A, Bouquet A, Maucourt M, Thomas MR, Torregrosa L.** The grapevine *Fleshless berry* mutation. A unique genotype to investigate differences between fleshy and nonfleshy fruit. *Plant Physiol* 2006b, 140:537-547.

**Fernandez L, Torregrosa L, Segura V, Bouquet A, Martinez-Zapater JM.** Transposon-induced gene activation as a mechanism generating cluster shape somatic variation in grapevine. *Plant J* 2010, 61:545-57.



**Fernandez L, Torregrosa L, Terrier N, Sreekantan L, Grimplet J, Davies C, Thomas M.R, Romieu C, Ageorges A.** Identification of genes associated with flesh morphogenesis during grapevine fruit development. *Plant Mol Biol* 2007, 63:307-323.

**Ferrandiz C, Pelaz S, Yanofsky MF.** Control of carpel and fruit development in Araidopsis. *Ann Rev Biochem* 1999, 68:321-354.

**Fischer B, Salakhutdinov I, Akkurt M, Eibach R, Edwards KJ, Töpfer R, Zyprian E.** Quantitative trait locus analysis of fungal disease resistance factors on a molecular map of grapevine. *Theoretical and applied genetics* 2004, 108 :501-515.

**Flint-Garcia SA, Thornsberry JM, Buckler ES.** Structure of linkage disequilibrium in plants. *Annu. Rev Plant Biol* 2003, 54: 357-374.

**Flint-Garcia SA, Thuillet AC, Yu JM, Pressoir G, Romero SM, Mitchell SE, Doebley J, Kresovich S, Goodman MM, Buckler ES.** Maize association population: a high-resolution platform for quantitative trait locus dissection. *Plant J* 2005, 44:105-1064.

**Fournier-Level A, Lacombe T, Le Cunff L, Boursiquot JM, This P.** Evolution of the VvMybA gene family, the major determinant of berry colour in cultivated grapevine (*Vitis vinifera* L.). *Heredity* 2010, 104:351-62.

**Fournier-Level A, LeCunff L, Gomez C, Doligez A, Ageorges A, Roux C, Bertrand Y, Souquet JM, Cheynier V This P.** Quantitative genetic bases of anthocyanin variation in grape (*Vitis vinifera* L. ssp. *sativa*) berry: a quantitative trait locus to quantitative trait nucleotide integrated study. *Genetics* 2009, 183:1127-1139.

**Franks T, Botta R, Thomas MR.** Chimerism in grapevines: implications for cultivar identity, ancestry and genetic improvement. *Theor Appl Genet* 2002, 104: 192-199.

**Frary A, Nesbitt TC, Frary A, Grandillo S, van der Knaap E, Cong B, Liu J, Meller J, Elber R, Alpert KB, Tanksley SD:** fw2.2: A Quantitative Trait Locus key to the evolution of tomato fruit size. *Science* 2000, 289:85-88.

**Fregoni M.** Origines de la vigne et de la viticulture. Contribution des peuples antiques. Ed Musumeci, Italy, 1991, 158p.

**Fridman E, Carrari F, liu Y-S, Fernie A, Zamir D:** Zooming in on a quantitative trait for for tomato yield using interspecific introgressions. *Science* 2004, 305:1786-1789.

**Fulton TM, Bucheli P, Voirol E, Lopez J, Pétiard V, Tanksley SD.** Quantitative trait loci (QTL) affecting sugars, organic acids, and other biochemical properties possibly contributing to flavor, identified in four advanced backcross populations of tomato. *Euphytica* 2002, 127:163-177.

**Galet P .**Dictionnaire encyclopédique des cépages. Ed Hachette, Paris, France, 2000, 1024p.

**Galet P.** Précis de viticulture. Ed Galet, Montpellier, France, 1993, 602p.

**Gao LZ, Zhang CH, Li DY, Pan DJ, Jia JZ, Dong YS.** Genetic diversity within *Oryza rufipogon* germplasms preserved in Chinese field gene banks of wild rice as revealed by microsatellite markers. *Biodiver Conserv* 2006, 15: 4059-4077.

**Gerlach W, Bedbrook J.** Cloning and characterization of ribosomal RNA genes from wheat and barley. *Nucl Acids Res* 1979, 7:1869-1885.

**Gillaspy G, Ben-David H, Gruissem W.** Fruits: a developmental perspective. *Plant Cell* 1993, 5:1439-1451.

**Glad C, Regnard JL, Querou Y, Brun O, Morot-Gaudry JF.** Flux and chemical composition of xylem exudates from chardonnay grapevines: temporal evolution and effet of recut. *Am J Enol Vitic* 1992, 4:275-282.

**Golzarian MR, Frick RA, Rajendran K, Berger B, Roy S, Tester M, Lun DS.** Accurate inference of shoot biomass from high-throughput images of cereal plants. *Plant Methods* 2011, 7:2.

**Gonzalez N, Gevaudant F, Hernould , Chevalier C, Mouras A.** The cell cycle-associated protein kinase WEE1 regulates cell size in relation to endoreduplication in developing tomato fruit. *The Plant Journal* 2007, 51: 642 – 655.

**Grandillo S, Ku HM, Tanksley SD.** Identifying loci responsible for natural variation in fruit size and shape in tomato. *Theor Appl Genet* 1999, 99:978-987.

**Granier C, Aguirrezabal L, Chenu K, Cookson SJ, Dauzat M, Hamard P, Thioux JJ, Rolland G, Bouchier-Combaud S, Lebaudy A, Muller B, Simonneau T, Tardieu F.** PHENOPSIS, an automated platform for reproducible phenotyping of plant responses to soil water deficit in *Arabidopsis thaliana* permitted the identification of an accession with low sensitivity to soil water deficit. *New Phytol* 2006, 169:623-35.

**Grassi F, De Mattia F, Zecca G, Sala F, Labra M.** Historical isolation and Quaternary range expansion of divergent lineages in wild grapevine. *Biological Journal of the Linnean Society* 2008, 95:611-619.

**Grassi F, et al.** Evidence of a secondary grapevine domestication centre detected by SSR analysis. *Theor Appl Genet* 2003, 107:1315-1320.

**Grattapaglia D, Bertolucci FLG, Penchel R, et al.** Genetic mapping of quantitative trait loci controlling growth and wood quality traits in *Eucalyptus grandis* using a maternal half-sib family and RAPD markers. *Genetics* 1996, 144:1205-1214.

**Grattapaglia D, Bertolucci FLG, Penchel R, et al.** Genetic mapping of quantitative trait loci controlling growth and wood quality traits in *Eucalyptus grandis* using a maternal half-sib family and RAPD markers. *Genetics* 1996, 144:1205-1214.

**Gross BL, Olsen KM.** Genetic perspectives on crop domestication. *Trends Plant Sci* 2010, 15:529-37.

**Gu Q, Ferrándiz C, Yanofsky MF, Martienssen R.** The FRUITFULL MADS-box gene mediates cell differentiation during Arabidopsis fruit development. *Development* 1998, 125:1509-17.

**Guo M, Rupe MA, Dieter JA, Zou J, Spielbauer D, Duncan KE, Howard RJ, Hou Z, Simmons CR.** Cell Number Regulator1 affects plant and organ size in maize: implications for crop yield enhancement and heterosis. *Plant Cell* 2010, 22:1057-73.

**Gupta P.K, Rustgir S, Kulwal PL.** Linkage disequilibrium and association studies in higher plants: Present status and future prospects. *Plant Molecular Biology* 2005, 57:461-485.

**Gupta V, Mathur S, Solanke AU, Sharma MK, Kumar R, Vyas S, Khurana P, Khurana JP, Tyagi AK, Sharma AK.** Genome analysis and genetic enhancement of tomato. *Crit Rev Biotechnol* 2009, 29:152-81.

**Hamblin MT, Buckler ES, Jannink JL.** Population genetics of genomics-based crop improvement methods. *Trends Genet* 2011, 27:98-106.

**Hamblin MT, Casa AM, Sun H, Murray SC, Paterson AH, Aquadro CF, Kresovich S.** Challenges of detecting directional selection after a bottleneck: Lessons from Sorghum bicolor. *Genetics* 2006, 173: 953-964.

**Hardie WJ, O'Brien TP, Jaudzems VG.** Morphology, anatomy and development of the pericarp after anthesis in grape, *Vitis vinifera* L. *Australian J Grape Wine Research* 1996, 2: 97-142.

**Harris JP, Kriedemann PE, Possingham JV.** Anatomical aspects of grape berry development. *Vitis* 1968, 7: 106-119.

**Hartl DL, Clark AG.** Principles of Population Genetics. Ed. Sinauer Associates, Sunderland, Massachusetts, USA, 1997, 565p.

**Hilu KW, Borsch T, Müller K, Soltis DE, Soltis PS, Savolainen V, Chase MW, Powell MP, Alice LA, Evans R, Sauquet H, Neinhuis C, Slotta TAB, Rohwer JG, Campbell CS, Chatrou LW.** Angiosperm phylogeny based on matK sequence information. *American Journal of Botany* 2003, 90:1758-1776.

**Hocquigny S, Pelsy F, Dumas V, Kindt S, Heloir MC, Merdinoglu D.** Diversification within grapevine cultivars goes through chimeric states. *Genome* 2004, 47:579-589.

**Hocquigny S.** La diversité génétique intra-variétale chez la vigne : caractérisation et origines. Thèse de l'Université de Strasbourg, France, 2003, 154p.

**Hoffmann S, Di Gaspero G, Kovács L, Howard S, Kiss E, Galbács Z, Testolin R, Kozma P.** Resistance to Erysiphe necator in the grapevine 'Kishmish vatkana' is controlled by a single locus through restriction of hyphal growth. *Theor Appl Genet* 2008, 116:427-38.

**Houel C, Bounon R, Chaïb J, Guichard C, Péros JP, Bacilieri R, Dereeper A, Canaguier A, Lacombe T, N'Diaye A, Le Paslier MC, Vernerey MS, Coriton O, Brunel D, This P,**

**Torregrosa L, Adam-Blondon AF.** Patterns of sequence polymorphism in the fleshless berry locus in cultivated and wild *Vitisvinifera* accessions. *BMC Plant Biol* 2010, 10:284.

**Houel C, Martin-Magniette M-L, Nicolas SD, Lacombe T, Le Cunff L, Franck D, Torregrosa L, Con  j  ro G, Lalet S, This P, Adam-Blondon A-F.** Genetic variability of the berry size in grapevine (*Vitis vinifera* L.) involves ovary and early berry development associated factors. *J Exp Bot* submitted.

**Howe KL, Chothia T, DurbinR.** GAZE: A Generic Framework for the Integration of Gene-Prediction Data by Dynamic Programming. *Genome Res* 2002, 12:1418-1427.

**Jack T.** Relearning our ABCs: new twists on an old model. *Trends Plant Sci* 2001, 6:310-6.

**Jackson DI and Coombe BG.** The growth of apricot fruit. I. Morphological changes during development and the effects of various tree factors. *Aust J Agri Res* 1966, 17:465-477.

**Jacobsen SE, Olszewski NE.** Mutations at the SPINDLY locus of *Arabidopsis* alter gibberellin signal transduction. *Plant Cell* 1993, 5:887-896.

**Jaillon O, Aury J-M, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C, Vezzi A, Legeai F, Hugueney P, Dasilva C, Horner D, Mica E, Jublot D, Poulain J, Bruyere C, Billault A, Segurens B, Gouyvenoux M, Ugarte E, Cattonaro F, Anthouard V, Vico V, Del Fabbro C, Alaux M, Di Gaspero G, Dumas V, Felice N, Paillard S, Juman I, Moroldo M, Scalabrin S, Canaguier A, Le Clainche I, Malacrida G, Durand E, Pesole G, Laucou V, Chatelet P, Merdinoglu D, Delledonne M, Pezzotti M, Lecharny A, Scarpelli C, Artiguenave F, P   E, Valle G, Morgante M, Caboche M, Adam-Blondon A-F, Weissenbach J, Qu  tier F, Wincker P.** The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 2007, 449:463-468.

**Jelesko JG, Jenkins SM, Rodriguez-Concepcion M, Gruissem W.** Regulation of tomato HMG1 during cell proliferation and growth. *Planta* 1999, 208:310-318.

**Khawale RA, Yerramilli V, Singh SK.** Molecular marker-assisted selection of in vitro chemical mutagen-induced grapevine mutants. *Curr Science* 2007, 92: 1056-1060.

**Kim GH, Yun HK, Choi CS, Park JH, Jung YJ, Park KS, Dane F, Kang KK.** Identification of AFLP and RAPD markers linked to anthracnose resistance in grapes and their conversion to SCAR markers. *Plant Breed* 2008, 127:418-423.

**Kim S, Plagnol V, Hu TT, Toomajian C, Clark RM, Ossowski S, Ecker JR, Weigel D, Nordborg M.** Recombination and linkage disequilibrium in *Arabidopsis thaliana*. *Nat Genet* 2007, 39:1151-5.

**Kim Y, Stephan W.** Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics* 2002, 160: 765-777.

**Kimura M.** Evolutionary rate at the molecular level. *Nature* 1968, 217: 624- 626.

**Klekowski EJ, Kazarinovafukshansky N, Mohr H.** Shoot apical meristems and mutation – stratified meristems and angiosperm evolution. *American Journal of Botany* 1985, 72:1788-1800.

**Kliewer WM.** Effect of high temperatures during the bloom-set period, ovule fertility, and berry growth of several grape cultivar. *Am J Enol Vitic* 1977, 28:215-222.

**Kobayashi T, Kato-Emori S, Tomita K, Ezura H.** Detection of 3-hydroxy-3-methylglutaryl-coenzyme A reductase protein Cm-HMGR during fruit development in melon (*Cucumis melo* L.). *TheorAppl Genet* 2002, 104:779-785.

**Konishi S, Izawa T, Lin SY, Ebana K, Fukuta Y, Sasaki T, Yano M.** An SNP caused loss of seed shattering during rice domestication. *Science*, 2006 312:1392-6.

**Konishi, S. et al.** An SNP caused loss of seed shattering during rice domestication. *Science* 2006, 312:1392-1396.

**Krivanek AF, Riaz S, Walker MA.** Identification and molecular mapping of PdR1, a primary resistance gene to Pierce's disease in Vitis. *Theor Appl Genet* 2006, 112:1125-1131.

**Kuksova VB, Piven NM, Gleba YY.** Somaclonal variation and in vitro induced mutagenesis in grapevine. *Plant Cell Tiss Organ Cult* 1997, 49:17-27.

**Labate JA, Robertson LD, Wu FN, Tanksley SD, Baldo AM.** EST, COSII, and arbitrary gene markers give similar estimates of nucleotide diversity in cultivated tomato (*Solanum lycopersicum* L.). *Theor Appl Genet* 2009, 118:1005-1014.

**Lacombe T, Laucou V, di Vecchi M, Bordenave L, Bourse T, Siret R, David J, Boursiquot J-M, Bronner A, Merdinoglu D, This P.** Contribution à la caractérisation et à la protection in situ des populations de Vitis vinifera L. ssp. silvestris (Gmelin) Hegi. Ed Les Actes du BRG, France, 2003, 4:381-404.

**Lahogue F, This P, Bouquet A.** Identification of a codominant marker linked to the seedlessness character in grapevine. *Theor Appl Genet* 1998, 97: 950-959.

**Lakso AN, Kliewer WM.** Physical properties of phosphoenolpyruvate carboxylase and malic enzyme in grape berries. *Am J Enol Vitic* 1975, 26:75-78.

**Lamoureux D, Bernole A, Le Clainche I, Tual S, Thareau V, Paillard S, Legeai F, Dossat C, Wincker P, Oswald M, Merdinoglu D, Vignault C, Delrot S, Caboche M, Chalhoub B, Adam-Blondon AF:** Anchoring of a large set of markers onto a BAC library for the development of a draft physical map of the grapevine genome. *Theor Appl Genet* 2006, 113: 344-356.

**Laucou V, Lacombe T, Dechesne F, Siret R, Bruno J-P, Dessup M, Dessup T, Ortigosa P, Parra P, Roux C, Santoni S, Vares D, Peros J-P, Boursiquot J-M, This P.** High throughput analysis of grape genetic diversity as a tool for germplasm collection management. *Theor Appl Genet* 2011, DOI 10.1007/s00122-010-1527-y.

**Lavee S, Nir G.** Grape. Handbook of fruit set and development. Ed Monselise SP, Florida, USA, 1986, p167-191.

**Le Cunff L, Fournier-Level A, Laucou V, Vezzulli S, Lacombe T, Adam-Blondon A-F, Boursiquot J-M, This P.** Construction of nested core collections to optimize the exploitation of natural diversity in *Vitis vinifera* L. subsp. *sativa*. *BMC Plant Biology* 2008, 2:8-31.

**Leflon M, Eber F, Letanneur JC, Chelysheva L, Coriton O, Huteau V, Ryder CD, Barker G, Jenczewski E, Chevre AM.** Pairing and recombination at meiosis of *Brassica rapa* (AA) x *Brassica napus* (AACC) hybrids. *Theor Appl Genet* 2006, 113:1467-1480.

**Lemaire-Chamley M, Petit J, Garcia V, Just D, Baldet P, Germain V, Fagard M, Mouassite M, Cheniclet C, Rothan C.** Changes in transcriptional profiles are associated with early fruit tissue specialization in tomato. *Plant Physiol* 2005, 139:750-769.

**Levadoux L.** Les Populations sauvages et cultivées de *Vitis vinifera* L. *Annales de l'Amélioration de Plantes* 1956, 6:59-118.

**Li CB, Zhou AL, Sang T.** Rice domestication by reducing shattering. *Science* 2006, 311:1936-1939.

**Libault M, Zhang XC, Govindarajulu M, Qiu J, Ong YT, Brechenmacher L, Berg RH, Hurley-Sommer A, Taylor CG, Stacey G.** A member of the highly conserved FWL (tomato FW2.2-like) gene family is essential for soybean nodule organogenesis. *Plant J* 2010, 62:852-64.

**Lijavetzky D, Cabezas J, Ibanez A, Rodriguez V, Martinez-Zapater JM.** choice of the candidate genes. High throughput SNP discovery and genotyping in grapevine (*V. vinifera* L.) by combining a re-sequencing approach and SNPlex technology. *BMC Genome* 2007, 8:424.

**Lijavetzky D, Ruiz-García L, Cabezas JA, De Andrés MT, Bravo G, Ibáñez A, Carreño J, Cabello F, Ibáñez J, Martínez-Zapater JM.** Molecular genetics of berry colour variation in table grape. *Mol Genet Genomics* 2006, 276:427-35.

**Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, et al.** Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* 2005, 438:803-819.

**Lippman ZB, Semel Y, Zamir D.** An integrated view of quantitative trait variation using tomato interspecific introgression lines. *Curr Opin Genet & Dev* 2007, 17:545-552.

**Lippman Z, Tanksley SD.** Dissecting the genetic pathway to extreme fruit size in tomato using a cross between the small-fruited wild species *L. pimpinellifolium* and *L. esculentum*, var. Giant Heirloom. *Genetics* 2001, 158:413-422.

**Liu J, Van Eck J, Cong B, Tanksley SD.** A new class of regulatory genes underlying the cause of pear-shaped tomato fruit. *Proc Natl Acad Sci USA* 2002, 99:13302-6.

**Liu JP, Cong B, Tanksley SD.** Generation and analysis of an artificial gene dosage series in tomato to study the mechanisms by which the cloned quantitative trait locus fw2.2 controls

fruit size. *Plant Physiol* 2003, 132:292-299.

**Lorenz DH, Eichhorn KW, Blei-holder H, Klose R, Meier H, Weber E.** Phänologische Entwicklungsstadien der Weinrebe (*Vitis vinifera* L. ssp. *vinifera*). *Vitic Enol Sci* 1994, 49:66-70.

**Lowe KM, Walker MA.** Genetic linkage map of the interspecific grape rootstock cross Ramsey (*Vitis champinii*) x Riparia Gloire (*Vitis riparia*). *Theor Appl Genet* 2006, 112: 1582-1592.

**Lozano R, Giménez E, Cara B, Capel J, Angosto T.** Genetic analysis of reproductive development in tomato. *Int J Dev Biol* 2009, 53:1635-1648.

**Lund ST, Peng FY, Nayar T, et al.** Gene expression analyses in individual grape (*Vitis vinifera* L.) berries during ripening initiation reveal that pigmentation intensity is a valid indicator of developmental staging within the cluster. *Plant Molecular biology* 2008, 68:301-315.

**MacArthur JW, Butler L.** Size inheritance and geometric growth processes in the tomato fruit. *Genetics* 1938, 23:253-268.

**Marguerit E, Boury C, Manicki A, Donnart M, Butterlin G, Némorin A, Wiedemann-Merdinoglu S, Merdinoglu D, Ollat N, Decroocq S.** Genetic dissection of sex determinism, inflorescence morphology and downy mildew resistance in grapevine. *Theor Appl Genet* 2009, 118:1261-1278.

**Marinval P.** Vigne sauvage et vigne cultivée dans le bassin méditerranéen - Emergence de la viticulture - Contribution archéo-botanique. Histoire du vin, une histoire de rites. Paris: OIV, 1997, p137-172.

**Martin DM, Aubourg S, Schouwey MB, Daviet L, Schalk M, Toub O, Lund ST, Bohlmann J.** Functional annotation, genome organization and phylogeny of the grapevine (*Vitis vinifera*) terpene synthase gene family based on genome assembly, FLcDNA cloning, and enzyme assays. *BMC Plant Biol* 2010, 10:226.

**Mather KA, Caicedo AL, Polato NR, Olsen KM, McCouch S, Purugganan MD.** The extent of linkage disequilibrium in rice (*Oryza sativa* L.). *Genetics* 2007, 177:2223-32.

**Mathieu-Rivet E, Gevaudant F, Sicard A, et al.** The functional analysis of the Anaphase Promoting Complex activator CCS52A high- lights the crucial role of endoreduplication for fruit growth in tomato. *The Plant Journal* 2010, 62: 727-741.

**Mccarthy MG.** Weight loss from ripening berries of Shiraz grapevines (*Vitis vinifera* L. cv Shiraz). *Aust J Grape Wine Res* 1999, 5:10-16.

**McGovern PE.** Ancient wine: the search for the origins of viniculture. Ed Princeton Univ Princeton, New Jersey, USA, 2003, 400p.

**Mejia N, Gebauer M, Muñoz L, Hewstone N, Muñoz C, Hinrichsen P.** Identification of QTL for seedlessness, berry size, and ripening date in a seedless x seedless progeny. *Am J Enol Vitic* 2007, 58:499-507.

**Mejia N, Soto B, Guerrero M, Casanueva X, Miccona M, Houel C, Le Cunff L, Boursiquot JM, Hinrichsen P, Adam-Blondon AF.** Molecular, genetic and transcriptional evidence for a role of *VvAGL11* in stenospermocarpic seedlessness in grapevine. *BMC plant biology*, 2011 in press.

**Menda N, Semel Y, Peled D, Eshed Y, Zamir D.** In silico screening of a saturated mutation library of tomato. *The Plant J* 2004, 38:861-872.

**Merdinoglu D, Wiedemann-Merdinoglu S, Coste P, Dumas V, Haetty S, Butterlin G, Greif C.** Genetic analysis of downy mildew resistance derived from *Muscadinia rotundifolia*. *Acta Hort* 2003, 603: 451-456.

**Miller JC, Tankley SD.** RFLP analysis of phylogenetic relationships and genetic variation in the genus *Lycopersicon*. *Theor Appl Genet* 1990, 80:437-448.

**Montes JM, Melchinger AE, Reif JC.** Novel throughput phenotyping platforms in plant genetic studies. *Trends Plant Sci* 2007, 12:433-6.

**Morgan T.** Random segregation versus coupling in Mendelian inheritance. *Science* 1911, 34:384.

**Mullins MG, Bouquet A, Williams LE.** Biology of the grapevine. Ed Cambridge University, USA, 1992, 239p.

**Myles S, Boyko AR, Owens CL, Brown PJ, Grassi F, Aradhya MK, Prins B, Reynolds A, Chia JM, Ware D, Bustamante CD, Buckler ES.** From the Cover: Genetic structure and domestication history of the grape. *Proc Natl Acad Sci* 2011, 108:3530-5.

**Myles S, Chia J-M, Hurwitz B, Simon C, Zhong GY, Buckler E, Ware D.** Rapid genomic characterization of the genus *Vitis*. *Plos One* 2010, 5:e8219.

**Myles S, Peiffer J, Brown PJ, Ersoz ES, Zhang Z, Costich DE, Buckler ES.** Association mapping: critical considerations shift from genotyping to experimental design. *Plant Cell* 2009, 21:2194-202.

**Nagarajan N, Pop M.** Sequencing and genome assembly using next-generation technologies. *Methods Mol Biol* 2010, 673:1-17.

**Narita JO, and Gruissem W.** Tomato hydroxyl-methyl-glutaryl-CoA reductase is required early in fruit development but not during ripening. *Plant Cell* 1989, 1:181-190.

**Neale DB, Savolainen O.** Association genetics of complex traits in conifers. *Trends in Plant Science* 2004, 9:325-330.

**Negrul AM.** Evolucija kuljturnyx form vinograda. *Doklady Akademii nauk SSSR* 1938, 8: 585-585.



**Nesbitt TC, Tanksley SD.** Comparative sequencing in the genus *Lycopersicon*. Implications for the evolution of fruit size in the domestication of cultivated tomatoes. *Genetics* 2002, 162:365-79.

**Nishimura T, Wada T, Yamamoto KT, Okada K.** The Arabidopsis STV1 protein, responsible for translation reinitiation, is required for auxin-mediated gynoecium patterning. *Plant Cell* 2005, 17:2940-53.

**Ojeda H, Deloir A, Carbonneau A.** Influence of water deficits on grape berry growth. *Vitis* 2001, 40:141-145.

**Ojeda H, Deloire A, Carbonneau A, Ageorges A, Romieu C.** Berry development of grapevines: Relations between the growth of berries and their DNA content indicate cell multiplication and enlargement. *Vitis* 1999, 38:145-150.

**Ollat N, Diakou-Verdin P, Carde JP, Barrieu F, Gaudillere JP, Moing A.** Grape berry development: A review. *J Int Sci Vigne Vin* 2002, 36:109-131.

**Olmo HP.** Bud mutation in the vinifera grape. 2. Sultanina gigas. *Proc Am Soc Hort Sci* 1935, 33:437- 439.

**Olmo HP.** Bud mutation in the vinifera grape: 1. Parthenocarpic sultanina. *Proc Am Soc Hort Sci* 1934, 31:118-121.

**Olmo HP.** Grapes: *Vitis*, *Muscadinia* (Vitaceae). In: Evolution of crop plants. Ed Simmonds NW, New york, USA, 1976, pp294-298.

**Olsen KM, Caicedo AL, Polato N, McClung A, McCouch S, Purugganan MD.** Selection under domestication: evidence for a sweep in the rice waxy genomic region. *Genetics* 2006, 173:975-83.

**Palaisa KA, Morgante M, Williams M, Rafalski A.** Contrasting effects of selection on sequence diversity and linkage disequilibrium at two phytoene synthase loci. *Plant Cell* 2003, 15:1795-1806.

**Palaisa KA, Morgante M, Williams M, Rafalski A.** Long-range patterns of diversity and linkage disequilibrium surrounding the maize Y1 gene are indicative of an asymmetric selective sweep. *PNAS* 2004, 101:9885-9890.

**Paran I, van der Knaap E.** Genetic and molecular regulation of fruit and plant domestication traits in tomato and pepper. *J Exp Bot* 2007, 58:3841-52.

**Park HS.** Le péricarpe des baies de raisin normales et millerandées : ontogenèse de la structure et évolution de quelques constituants biochimiques, notamment les tanins. Thèse de l'Université de Bordeaux II, France, 1995, 208p.

**Paterson AH.** What has QTL mapping taught us about plant domestication? *New Phytologist* 2002, 154:591-608.

**Pauquet J, Bouquet A, This P, Adam-Blondon AF.** Establishment of a local map of AFLP markers around the powdery mildew resistance gene Run1 in grapevine and assessment of their usefulness for marker assisted selection. *Theor Appl Genet* 2001, 103:1201-1210.

**Pelaz S, Ditta GS, Baumann E, Wisman E, Yanofsky MF.** B and C floral organ identity functions require SEPALLATA MADS-box genes. *Nature* 2000, 405:200-203.

**Pelaz S, Gustafson-Brown C, Kohalmi SE, Crosby WL, Yanofsky MF.** APETALA1 and SEPALLATA3 interact to promote flower development. *Plant J* 2001, 26:385-394.

**Pelsy F, Hocquigny S, Moncada X, Barbeau G, Forget D, Hinrichsen P, Merdinoglu D.** An extensive study of the genetic diversity within seven French wine grape variety collections. *Theor Appl Genet* 2010, 120:1219-31.

**Pelsy F.** Molecular and cellular mechanisms of diversity within grapevine varieties. *Heredity* 2010, 104:331-40.

**Pelsy F.** Untranslated leader region polymorphism of Tvv1, a retrotransposon family, is a novel marker useful for analyzing genetic diversity and relatedness in the genus *Vitis*. *Theor Appl Genet* 2007, 116:15-27.

**Peng JH, Bai Y, Haley SD, Lapitan NLV.** Microsatellite-based molecular diversity of bread wheat germplasm and association mapping of wheat resistance to the Russian wheat aphid. *Genetica* 2009, 135:95-122.

**Philippe R, Courtois B, McNally KL, Mournet P, El-Malki R, Le Paslier MC, Fabre D, Billot C, Brunel D, Glaszmann JC, This D.** Structure, allelic diversity and selection of Asr genes, candidate for drought tolerance, in *Oryza sativa* L. and wild relatives. *Theor Appl Genet* 2010, 121:769-787.

**Picken A.** A review of pollination and fruit set in the tomato (*Lycopersicon esculentum* Mill). *J Horticult Sci* 1984, 59:1-13.

**Pouget R.** Histoire de la lutte contre le Phylloxéra de la vigne en France. OIV, INRA, Paris, 1990, p156.

**Pouget R.** *Vitis vinifera*, histoire et évolution. In *La Vigne et le Vin*. Ed La Manufacture et la Cité des sciences et de l'industrie, Paris, France, 1988, pp15-25.

**Poupin MJ, Federici F, Medina C, Mattis JT, Timmermann T, Arce-Johnson P.** Isolation of the three grape sub-lineages of B-class MADS-box TM6, PISTILLATA and APETALA3 genes which are differentially expressed during flower and fruit development. *Gene* 2007, 404:10-24.

**Pratt C.** Reproductive anatomy in cultivated grapes – A review. *Am J Enol Vitic* 1971, 22: 92-109.

**Pritchard JK, Stephens M, Donnelly P.** Inference of population structure using multilocus genotype data. *Genetics* 2000, 155:945-959.

**Puchta H.** The repair of double-strand breaks in plants: mechanisms and consequences for genome evolution. *J Exp Bot* 2005, 56:1-14.

**Purugganan MD, Fuller DQ.** The nature of selection during plant domestication. *Nature* 2009, 457:843-848.

**Rafalski A, Morgante M.** Corn and humans: recombination and linkage disequilibrium in two genomes of similar size. *Trends Genet* 2004, 20:103-11.

**Rafalski JA.** Plant genomics: present state and a perspective on future developments. *Brief Funct Genomic Proteomic* 2002, 1:80-94.

**Ranc N, Muños S, Santoni S, Causse M.** A clarified position for *Solanum lycopersicum* var. *cerasiforme* in the evolutionary history of tomatoes (solanaceae). *BMC Plant Biol* 2008, 8:130.

**Ranc N.** Analyse du polymorphisme moléculaire de gènes de composantes de la qualité des fruits dans les ressources génétiques sauvages et cultivées de tomate; recherche d'association genes/QTL. Thèse de l'Université de Montpellier, France, 2010, 273p.

**Reid KE, Olsson N, Schlosser J, Peng F, Lund ST.** An optimized grapevine RNA isolation procedure and statistical determination of reference genes for real-time RT-PCR during berry development. *BMC Plant Biology* 2006, 14:6-27.

**Riaz S, Dangl GS, Edwards KJ, Meredith CP.** A microsatellite marker based framework linkage map of *Vitis vinifera* L. *Theor Appl Genet* 2004, 108:864-872.

**Riaz S, Krivanek AF, Xu K, Walker MA.** Refined mapping of the pierces disease resistance locus, PdR1, and sex on the extended genetic map of *Vitis rupestris* x *V. arizonica*. *Theor Appl Genet* 2006, 113: 1317-1329.

**Roger AR, Huff C.** Linkage disequilibrium between loci with unknown phase. *Genetics* 2009, 182:839-844.

**Rounsley S, Ditta G, Yanofsky M.** Diverse roles for MADS box genes in Arabidopsis development. *Plant Cell* 1995, 7:1259-1269.

**Royer C.** Mouvement historiques de la vigne dans le monde. In *La Vigne et le Vin*. Ed La Manufacture et la Cité des sciences et de l'industrie, Paris, France, 1988, pp15-25.

**Rozen S, Skaletsky HJ.** Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* 2000, 132:365-86.

**Salmaso M, Faes G, Segala C, Stefanini M, Salakhutdinov L, Zyprian E, Toepfer R, Grando MS, Velasco R.** Genome diversity and gene haplotypes in the grapevine (*V. vinifera* L.), as revealed by single nucleotide polymorphisms. *Mol Breed* 2004, 14:385-395.

**Salmaso M, Malacarne G, Troggio M, Faes G, Stefanini M, Grando MS, Velasco R.** A grapevine (*Vitis vinifera* L.) genetic map integrating the position of 139 expressed genes. *Theor Appl Genet* 2008, 116:1129-1143.

**Samson F, Brunaud V, Duchene S, et al.** FLAGdb(++): a database for the functional analysis of the Arabidopsis genome. *Nucleic acids research* 2004, 32:347-350.

**Savolainen V, Fay MF, Albach DC, Backlund A, van der Bank, Cameron KM, Johnson SA, Lledó MD, Pintaud J-C, Powell M, Sheahan MC, Soltis DE, Soltis PS, Weston P, Whitten WM, Wurdack KJ, Chase MW.** Phylogeny of the eudicots: a nearly complete familial analysis based on *rbcl* gene sequences. *Kew Bulletin* 2000, 55:257-309.

**Schmid KJ, Sorensen TR, Stracke R, Torjek O, Altmann T, Mitchell- Olds T, Weisshaar B.** Large-scale identification and analysis of genome-wide single-nucleotide polymorphisms for mapping in *Arabidopsis thaliana*. *Genome Res* 2003, 13:1250-1257.

**Schwarz-Sommer Z, Huijser P, Nacken W, Saedler H, Sommer H.** Genetic control of flower development by homeotic genes in *Antirrhinum majus*. *Science* 1990, 250: 931-936.

**Scienza A, Miravalle R, Visai C, et al.** Relationships between seed number, gibberellin and abscisic-acid levels and ripening in Cabernet Sauvignon grape berries. *Vitis* 1978, 17:361-368.

**Scienza A, Protti A, Conca E, Romano F.** Diffusione e caratteristiche della *Vitis vinifera silvestris* Gmelin in Italia. In 4<sup>3</sup> Simposio Internazionale di Genetica della Vite, Ed Ridvede Vignevini, Verona, Italia, 1885, p85.

**Sefc KM, Lopes MS, Lefort F, Botta R, Roubelakis-Angelakis KA., Ibanez J, Pejic I, Wagner HW, Glössl J, Steinkeller H.** Microsatellite variability in grapevine cultivars from different European regions and evaluation of assignment testing to assess the geographic origin of cultivars. *Theor Appl Genet* 2000, 100:498-505.

**Silacci MW, Morrison JC.** Changes in pectin content of Cabernet Sauvignon grape berries during maturation. *Am J Enol Vitic* 1990, 41:111-115.

**Silverstone AL, Tseng TS, Swain SM, Dill A, Jeong SY, Olszewski NE, Sun TP.** Functional analysis of SPINDLY in gibberellin signaling in *Arabidopsis*. *Plant Physiol* 2007, 143:987-1000.

**Slatkin M.** Linkage disequilibrium understanding the evolutionary past and mapping the medical future. *Nat Rev Genet* 2008, 9:477-485.

**Smartt J, Simmonds NW.** Evolution of Crop Plants. In Scientific and Technical. Ed Blackwell Science, London, UK, 1995, 496p.

**Snoussi H, et al.** Genetic relationship among cultivated and wild grapevine accessions from Tunisia. *Genome* 2004, 47:1211-219.

**Snyder E, Harmon FN.** Three mutations of *Vitis vinifera*. *Proc Amer Soc Hort Sci* 1935, 33: 435-436.

**Somers DJ, Kirkpatrick R, Moniwa M, Walsh A.** Mining single-nucleotide polymorphisms from hexaploid wheat ESTs. *Genome* 2003, 46: 431-437.

**Sreekantan L, Torregrosa L, Fernández L, Thomas MR.** VvMADS9, a class B MADS-box gene involved in grapevine flowering, shows different expression patterns in mutants with abnormal petal and stamen structures. *Funct Plant Biol* 2006, 33:877-886.

**Staden R.** The Staden Sequence Analysis Package. *Molecular Biotechnology* 1996, 5:233-241.

**Stich B, Maurer HP, Melchinger AE, Frisch M, Heckenberger M, Rouppe van der Voort J et al.** Comparison of linkage disequilibrium in elite European maize inbred lines using AFLP and SSR markers. *Mol Breeding* 2006, 17: 217–226.

**Suzuki M, Kamide Y, Nagata N, Seki H, Ohyama K, Kato H, Masuda K, Sato S, Kato T, Tabata S, Yoshida S, Muranaka T.** Loss of function of 3-hydroxy-3-methylglutaryl coenzyme A reductase 1 (HMG1) in Arabidopsis leads to dwarfing, early senescence and male sterility, and reduced sterol levels. *Plant J* 2004, 37:750-61.

**Sweeney MT, Thomson MJ, Cho YG, Park YJ, Williamson SH, Bustamante CD, McCouch SR.** Global dissemination of a single mutation conferring white pericarp in rice. *PLoS Genet* 2007, 3:e133.

**Tajima F.** Evolutionary relationship of DNA sequences in finite populations. *Genetics* 1983, 105: 437-460.

**Tajima F:** Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 1989, 123:585-595.

**Takahashi M, Matsuda F, Margetic N, Lathrop M.** Automated identification of single nucleotide polymorphisms from sequencing data. *Journal of Bioinformatics and Computational Biology* 2003, 1:253-265.

**Takeuchi T, Watanabe Y, Takano-Shimizu T, Kondo S.** Roles of jumonji and jumonji family genes in chromatin regulation and development. *Dev Dyn* 2006, 235:2449-59.

**Tamaru H.** Confining euchromatin/heterochromatin territory: jumonji crosses the line. *Genes Dev* 2010, 24:1465-78.

**Tanksley SD, Grandillo S, Fulton TM, et al.** Advanced backcross QTL analysis in a cross between an elite processing line of tomato and its wild relative *L-pimpinellifolium*. *Theoretical and applied genetics* 1996, 92: 213-224.

**Tanksley SD, McCouch.** Seed Banks and Molecular Maps: Unlocking Genetic Potential from the Wild. *Science* 1997, 277:1063-6.

**Tanksley SD.** The genetic, developmental, and molecular bases of fruit size and shape variation in tomato. *Plant Cell* 2004, 16:181-189.

**Tenaillon MI, Sawkins MC, Long AD, Gaut RL, Doebley JF, Gaut BS.** Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). *Proc Natl Acad Sci USA* 2001, 98: 9161-9166.

**Tenaillon MI, U'Ren J, Tenaillon O, Gaut BS.** Selection versus demography: a multilocus investigation of the domestication process in maize. *Mol Biol Evol* 2004, 21:1214-25.

**Terral J-F, Tabard E, Bouby L, Ivorra S, Pastor T, Figueiral I, Picq S, Chevance J-B, Jung C, Fabre L, Tardy C, Compan M, Bacilieri R, Lacombe T, This P.** Evolution and history of grapevine (*Vitis vinifera*) under domestication: new morphometric perspectives to understand seed domestication syndrome and reveal origins of ancient European cultivars. *Annals of Botany* 2010, 105:443-55.

**Terrier N, Glissant D, Grimplet J, Barrieu F, Abbal P, Couture C, Ageorges A, Atanassova R, Léon C, Renaudin JP, Dédaldéchamp F, Romieu C, Delrot S, Hamdi S.** Isogene specific oligo arrays reveal multifaceted changes in gene expression during grape berry (*Vitis vinifera* L.) development. *Planta* 2005, 222:832-847.

**Terrier N, Sauvage FX, Romieu C.** Absence de crise respiratoire, induction de l'activité alcool déshydrogénase et diminution de l'acide vacuolaire lors de la maturation du raisin. Ed A. Lonvaud-Fumel, Lavoisier, Paris, France, 1996. p24-28.

**Tesniere C, Pradal M, El-Kereamy A, Torregrosa L, Chatelet P, Roustan JP, Chervin C.** Involvement of ethylene signalling in a non-climacteric fruit: new elements regarding the regulation of ADH expression in grapevine. *J Exp Bot* 2004, 55:2235-40.

**Theissen G, Melzer R.** Molecular mechanisms underlying origin and diversification of the angiosperm flower. *Annals of botany* 2007, 100:603-619.

**Theissen G.** Development of floral organ identity: stories from the MADS house. *Curr Opin Plant Biol* 2001, 4:75-85.

**This P, et al.** Caractérisation de la diversité d'une population de vignes sauvages du Pic Saint-Loup (Hérault) et relations avec le compartiment cultivé. *Genet Selec Evol* 2001, 33: 289-304.

**This P, Lacombe T, Cadle-Davidson M, Owens CL.** Wine grape (*V. vinifera* L.) color associates with allelic variation in the domestication gene *VvmybA1*. *Theor Appl Genet* 2007, 114:723-730.

**This P, Lacombe T, Thomas MR.** Historical origins and genetic diversity of wine grapes. *Trends Genet* 2006, 22:511-519.

**Thomas MR, Cain P, Scott NS.** DNA Typing of Grapevines. A universal methodology and database for describing cultivars and evaluating genetic relatedness. *Plant Mol Biol* 1994, 25:939-949.

**Thompson MM, Olmo HP.** Cytohistological studies of cytochimeric and tetraploid grapes. *Am J Bot* 1963, 50: 901-906.

**Troggio M, Malacarne G, Coppola G, Segala C, Cartwright DA, Pindo M, Stefanini M, Mank R, Moroldo M, Morgante M, Grando MS, Velasco R.** A dense single-nucleotide polymorphism-based genetic linkage map of grapevine (*Vitis vinifera* L.). Anchoring Pinot Noir Bacterial Artificial Chromosome Contigs. *Genetics* 2007, 176:2637-2650.

**Tsukaya H.** Organ shape and size: a lesson from studies of leaf morphogenesis. *Curr Opin Plant Biol* 2003, 6:57-62.

**Van der Knaap E, Tanksley SD.** The making of a bell pepper-shaped tomato fruit: Identification of loci controlling fruit morphology in Yellow Stuffer tomato. *Theor Appl Genet* 2003, 107:139-147.

**Velasco R, Zharkikh A, Troggio M, Cartwright DA, Cestaro A, Pruss D, Pindo M, Fitzgerald LM, Vezzulli S, Reid J, Malacarne G, Iliev D, Coppola G, Wardell B, Micheletti D, Macalma T, Facci M, Mitchell JT, Perazzolli M, Eldredge G, Gatto P, Oyzerski R, Moretto M, Gutin N, Stefanini M, Chen Y, Segala C, Davenport C, Demattè L, Mraz A, Battilana J, Stormo K, Costa F, Tao Q, Si-Ammour A, Harkins T, Lackey A, Perbost C, Taillon B, Stella A, Solovyev V, Fawcett JA, Sterck L, Vandepoele K, Grando SM, Toppo S, Moser C, Lanchbury J, Bogden R, Skolnick M, Sgaramella V, Bhatnagar SK, Fontana P, Gutin A, Van de Peer Y, Salamini F, Viola R.** A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. *PLoS ONE* 2007, 2:e1326.

**Vezzulli S, Micheletti D, Riaz S, Pindo M, Viola R, This P, Walker MA, Troggio M, Ching AV, Caldwell KS, Jung M, Dolan M, Smith OS, Tingey S, Morgante M, Rafalski AJ.** SNP frequency, haplotype structure, linkage disequilibrium in elite maize inbred lines. *BMC Genet* 2002, 3:19.

**Vezzulli S, Micheletti D, Riaz S, Pindo M, Viola R, This P, Walker MA, Troggio M, Velasco R.** A SNP transferability survey within the genus *Vitis*. *BMC Plant Biology* 2008, 16:8-128.

**Vigouroux Y, Mitchell S, Matsuoka Y, Hamblin M, Kresovich S, Smith JSC, Jaqueth J, Smith OS, Doebley.** An analysis of genetic diversity across the maize genome using microsatellites. *Genetics* 2005, 169:1617-1630.

**Vouillamoz JF, Grando MS.** Genealogy of wine grape cultivars: Pinot is related to Syrah. *Heredity* 2006, 97:102-10.

**Walker AR, Lee E, Robinson SP.** Two new grape cultivars, bud sports of Cabernet Sauvignon bearing pale-coloured berries, are the result of deletion of two regulatory genes of the berry colour locus. *Plant Mol Biol* 2006, 62:623-635.

**Wang, H. et al.** The origin of the naked grains of maize. *Nature* 2005, 436:714-719 .

**Wang, RL, Stec A, Hey J, Lukens L, Doebley J.** The limits of selection during maize domestication. *Nature* 1999, 398:236-239.

**Watterson GA.** On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology* 1975, 7:188-193.

**Weaver R.J., Pool R.M.** Relation of seedness and ringing to gibberellin-like activity in berries of *Vitis vinifera*. *Plant Physiol* 1965, 40:770-776.

**Weeden NF.** Approaches to Mapping in Horticultural Crops. In PM Gresshoff. Ed Plant genome analysis, Boca Raton, USA, 1994, pp57-68.

**Weigel D, Meyerowitz EM.** The ABCs of floral homeotic genes. *Cell* 1994, 78:203-209.

**Welter LJ, Göktürk-Baydar N, Akkurt M, Maul E, Eibach R, Toepfer R, Zyprian E.** Genetic mapping and localization of quantitative trait loci affecting fungal disease resistance and leaf morphology in grapevine (*Vitis vinifera* L). *Mole Breed* 2007, 20:359-374.

**Winkler AJ, Cook JA, Kliewer WM, Lider LA.** General viticulture. Ed University of California, Los Angeles, USA, 1974, 710p.

**Wright SI, Bi IV, Schroeder SG, Yamasaki M, Doebley JF, McMullen MD, Gaut BS.** The effects of artificial selection of the maize genome. *Science* 2005, 308: 1310-1314.

**Xiao H, Jiang N, Schaffner E, Stockinger EJ, van der Knaap E.** A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. *Science* 2008, 319:1527-3.

**Xu K, Riaz S, Roncoroni NC, Jin Y, Hu R, Zhou R, Walker MA.** Genetic and QTL analysis of resistance to *Xiphinema* index in a grapevine cross. *Theor Appl Genet* 2008, 116:305-311.

**Yakushiji H, Kobayashi S, Goto-Yamamoto N, Tae Jeong S, Sueta T, Mitani N, Azuma A.** A skin color mutation of grapevine, from black-skinned Pinot Noir to white-skinned Pinot Blanc, is caused by deletion of the functional *VvmybA1* allele. *Biosci Biotechnol Biochem* 2006, 70:1506-8.

**Yakushiji H, Kobayashi S, Goto-Yamamoto N, Tae Jeong S, Sueta T, Mitani N, Azuma A.** A skin color mutation of grapevine, from black-skinned Pinot Noir to white-skinned Pinot Blanc, is caused by deletion of the functional *VvmybA1* allele. *Biosci Biotechnol Biochem* 2006, 70:1506-8.

**Yamane K, Tateishi K, Klose RJ, Fang J, Fabrizio LA, Erdjument-Bromage H, Taylor-Papadimitriou J, Tempst P, Zhang Y.** PLU-1 is an H3K4 demethylase involved in transcriptional repression and breast cancer cell proliferation. *Mol Cell* 2007, 25:801-12.

**Yamasaki M, Wright SI, McMullen MD.** Genomic screening for artificial selection during domestication and improvement in maize. *Ann Bot* 2007, 100:967-73.

**Yang Z, Oosterom EJ, Jordan DR, Hammer GL.** Pre-anthesis ovary development determines genotypic differences in potential kernel weight in sorghum. *J Exp Bot* 2009, 60:1399-1408.

**Yu J, Holland JB, McMullen MD, Buckler ES.** Genetic design and statistical power of nested association mapping in maize. *Genetics* 2008, 178:539-51.

**Yu JM, Buckler ES.** Genetic association mapping and genome organization of maize. *Curr Opin Biotechnol* 2006, 17:155-160.



**Zamboni A, Carli MD, Guzzo F, Stocchero M, Zenoni S, Ferrarini A, Tononi P, Toffali K, Desiderio A, Lilley KS, Pè ME, Benvenuto E, Delledonne M, Pezzotti M.** Identification of putative stage-specific grapevine berry biomarkers and omics data integration into networks. *Plant Physiology* 2010, 154:1439-59.

**Zhang J, Hausmann L, Eibach R, Welter LJ, Töpfer R, Zyprian E.** A framework map of grapevine V3125 (*Vitis vinifera* Schiava grossa x Riesling ) x rootstock cultivar Börner (*Vitis riparia* x *Vitis cinerea*) to localize genetic determinants of phylloxera root resistance. *Theor Appl Genet* 2009, 119:1039-1051.

**Zhao K, Aranzana MJ, Kim S, Lister C, Shindo C, Tang C, Toomajian C, Zheng H, Dean C, Marjoram P, Nordborg M.** Arabidopsis example of association mapping in structured samples. *PLoS Genet* 2007, 3:e4.

**Zhu C, M Gore, et al.** Status and prospects of association mapping in plants. *The Plant Genome* 2008, 1:5-20.

**Zhu YL, Song QJ, Hyten DL, Van Tassell CP, Matukumalli LK, Grimm DR, Hyatt SM, Fickus EW, Young ND, Cregan PB.** Single nucleotide polymorphisms in soybean. *Genetics* 2003, 163: 1123-1134.



## **8. Annexes**



**Annexe 1:** Figures et tableaux supplémentaires de  
l'article: *Genetic variability of the berry size in grapevine  
(Vitis vinifera L.) involves ovary and early berry development  
associated factors* (Chapitre 3)

Table S1. Plant material.

Name	Accession number	Sub-population <sup>§</sup>	Cluster group <sup>¶</sup>	Year of phenotyping
Abouhou	1992Mtp1	TE	—	—
Achlamiche	2991Mtp1	TE	D	2008
Aetonychi lefko	2281Mtp1	TE	NA	2008
Affenthaler	1583Mtp1	WW	A	2008
Ag isioum	1563Mtp1	TE	B	2009
Ag Kiourdach faux (Collection Kichinev)	2842Mtp1	TE	B	2008
Ahmar Derani	848Mtp1	TE	D	2008
Ahmeh Sal apyrène	2982Mtp1	TE	NA	2008
Ak ouzioum tagapskii*	2897Mtp1	—	—	—
Akiki	626Mtp1	TE	D	2008
Alba imputotato	44Mtp1	WE	—	—
Albaranzeuli bianco	2342Mtp1	WE	—	—
Alexandroouli	2500Mtp1	TE	A	2008
Alfrocheiro preto	50Mtp1	WW	A	2008
Altesse	167Mtp1	WW	A	2008
Alyarelhão <sup>§</sup>	1481Mtp1	—	A	2008
Angoor Kalan	0Mtp29	TE	C	2008
Ansonica	1307Mtp1	WE	—	—
Aouillat	403Mtp1	WW	B	2008
Araklinos*	1805Mtp1	WE	A	2008
Arakseni tcherni	2503Mtp1	TE	B	2008
Arbane	219Mtp1	WW	A	2008
Argant	261Mtp1	WE	NA	2008
Arinarnoa	2747Mtp1	WW	A	2008
Arinto tinto (Collection Soares Franco)	0Mtp36	WW	B	2008
Aromiesling	601Mtp1	WW	A	2008
Arvine	1837Mtp1	WW	A	2008
Assoued kere	1695Mtp1	TE	C	2008
Asswad Abou Khislé	2995Mtp1	TE	C	2008
Assyl kara*	2505Mtp1	TE	A	2008
Attiki	3005Mtp1	TE	C	2008
Autumn seedless	2893Mtp1	TE	C	2008
Avarengo	2218Mtp1	WW	NA	2008
Avgoustiatis	0Mtp64	WE	NA	2008
Azizi el Jaïa	2003Mtp1	TE	C	2009
B 40-97 (Ramming)	0Mtp1676	TE	B	2008
Bacchus = Geilweilerhof 33-29-133	2725Mtp1	WW	A	2008
Balbut bijeli	1888Mtp1	WE	B	2008
Balsamina (Collection Oberlin)	3009Mtp1	WW	B	2008
Baresana = Korithi aspro	1247Mtp1	TE	—	—
Barlinka	2225Mtp5	TE	D	2008-2009
Baroque	380Mtp1	WW	B	2008
Baserri	2890Mtp1	WW	B	2008-2009
Basicata	2298Mtp1	WE	A	2009
Belle Denise	0Mtp89	WW	A	2008
Bellone	1338Mtp1	WE	B	2008
Berbecel <sup>§</sup>	1573Mtp1	—	A	2008
Beregi rozsas	1629Mtp1	WE	A	2008
Bianco d'Alessano	2043Mtp1	WE	B	2008
Big Perlon	0Mtp1650	TE	—	—
Blanc Auba	344Mtp1	WW	B	2008-2009
Blanchier	2886Mtp1	WW	A	2008
Blank blauer	1631Mtp1	WE	B	2008
Bogazkere	2104Mtp1	TE	—	—
Bonamico	1354Mtp2	WE	B	2008
Bou Khanzir noir	1909Mtp1	TE	C	2008
Breider 5-6	2327Mtp1	WW	—	—
C 50-171 (Ramming)	0Mtp166	TE	—	—
Cabernet franc*	324Mtp1	—	NA	2009
Cacaboué	2070Mtp1	WW	A	2008
Cahours	0Mtp171	WW	B	2008
Calitor <sup>§</sup>	77Mtp1	—	B	2008
Camaraou noir	411Mtp1	WW	A	2008
Canella	0Mtp1758	WE	—	—
Caprignone	0Mtp1761	WE	—	—
Carcajolo	1531Mtp1	WW	B	2008
Carignan	18Mtp1	WE	B	2008-2009
Catarratto bianco	1303Mtp1	WE	NA	2008
Centennial seedless	2874Mtp1	TE	C	2008
César*	225Mtp2	—	NA	2009
Chahnani	1744Mtp1	TE	C	2008
Chami abiad	0Mtp220	TE	D	2008
Chaouch blanc	1673Mtp1	TE	D	2008
Chaptal	791Mtp1	TE	C	2008
Chasselas	585Mtp1	WW	B	2008
Chatus	129Mtp1	WW	A	2008
Chirai obak*	1186Mtp1	TE	B	2008
Chouchillon*	192Mtp1	—	A	2009
Clairette	26Mtp1	WW	A	2008
Claverie coulard	0Mtp1656	WW	B	2008-2009
Coarna alba	749Mtp1	WE	B	2008-2009
Cococciola	2348Mtp1	WE	—	—
Coda di volpe bianca	1261Mtp1	WE	NA	2008
Codivarta	2972Mtp1	WE	NA	2009
Corason de Kabritte blanc	0Mtp270	WW	C	2008
Corbeau	157Mtp1	WW	B	2008

Name	Accession number	Sub-population <sup>s</sup>	Cluster group <sup>†</sup>	Year of phenotyping
Courbu	413Mtp1	WW	A	2008
Crimposie	2243Mtp1	WE	C	2008
Dabouki	735Mtp1	TE	D	2008-2009
Danugue	666Mtp1	TE	C	2008
Darkaia noir	629Mtp1	TE	B	2008-2009
Dili kaftar	0Mtp300	TE		
Dimrit noir (Izmir)	2107Mtp1	WE	B	2008
Dinka zöld	1628Mtp1	WE		
Domina = Geilweilerhof 4-25-7	0Mtp313	WW	A	2008
Donzelinho branco	2892Mtp1	WW		
Doppel Augen	0Mtp318	TE		
Dureza	139Mtp1	WW	A	2008
Emerald seedless	2708Mtp1	TE	C	2008
Espadeiro tinto*	1498Mtp1			
Fer <sup>s</sup>	353Mtp1		A	2008
Feteasca regala	2247Mtp1	WE	A	2008
Fokiano	1797Mtp1	WE	NA	2009
Fortunato	0Mtp1733	TE		
Frankenthal rouge foncé	632Mtp1	TE	B	2008
Freisa	1245Mtp1	WW	A	2009
Fumin	2371Mtp1	WW	B	2008
Galbena de Odobesti	2136Mtp1	WE	B	2008
Galbena uriasa	2472Mtp1	WE	B	2008
Galotta	0Mtp1747	WW	A	2008-2009
Gantziandan	2902Mtp1	TE	B	2008
Garbanega faux (Istituto San Michele)	0Mtp408	TE	NA	2009
Garrido macho	1972Mtp1	TE	C	2008
Gascon	226Mtp1	WW	B	2008
Genk Uzum	789Mtp1	WE	B	2008
Gharbi	0Mtp416	TE	B	2008
Gheméra	329Mtp1	WE	C	2008
Gibert	372Mtp1	WW	B	2008
Gouais blanc	210Mtp1	WE	B	2008
Graeco	0Mtp440	TE		
Grain	0Mtp1513	WW	A	2008
Greco bianco	2373Mtp1	WE	B	2009
Gros Bourgogne	80Mtp1	WE	B	2008
Gros Cabernet	1232Mtp1	WW	A	2008
Grosse Méridle	0Mtp537	WW	B	2008
Hadari	1893Mtp1	TE		
Hagnos Zöld	0Mtp469	WE	B	2008
Hentakilo <sup>s</sup>	743Mtp1		C	2008
Heroldrebe = We S 130	2745Mtp1	WE	A	2008
Heunisch schwarz	1576Mtp1	WW	A	2008
Humagne blanc	2066Mtp1	WW	A	2008
Hunisa	635Mtp1	TE	D	2008-2009
Ichkimar	0Mtp586	TE	D	2008
Impigno	2057Mtp1	WE	B	2008
Joubertin	154Mtp1	WW	A	2008
July Muscat	2282Mtp1	TE	C	2008
Kadarum = Babic (Collection Split)	1644Mtp1	WE	B	2008
Kaisermuskat	613Mtp1	WW	A	2008
Kakotryghis	0Mtp562	WE	A	2008
Kapistroni tétri hermaphrodite (Coll. Kichinev)*	0Mtp567		B	2009
Kara oglan faux	0Mtp569	WE		
Karitsiotis	1784Mtp1	WE	B	2008
Kasoufi de la Bekaa	1715Mtp1	TE	B	2008
Katta-kourgan*	556Mtp1	TE		
Khalili belyi	657Mtp1	TE	NA	2008
Khalt	2013Mtp1	TE	B	2008
Khikhvi	0Mtp581	TE	A	2008
Khoussaïné blanc	0Mtp1614	TE		
Kichmich tcherni*	0Mtp583	TE	B	2008
Kizil sapak faux	746Mtp1	TE	B	2008
Kolliniatiko	0Mtp603	WE	B	2008
Kolontar	744Mtp1	WE	B	2008
Kolossié gebirgig	0Mtp604	TE		
Korza erevani	0Mtp610	TE	NA	2008
Koutlasky belyi	0Mtp1474	TE		
Kövidinka	1578Mtp1	WE	A	2008
Koz ouzioum	2635Mtp1	TE	B	2008
Labrusco	2953Mtp1	WW	A	2009
Lagrein	1287Mtp1	WW	A	2008
Lambrusco del Caset	0Mtp633	WW	A	2008
Lambrusco Marani	2349Mtp1	WW	A	2008
Lambrusque Grésienne 1*	8500Mtp124			
Lameiro*	0Mtp636	WW	A	2008
Landroter = Cornalin du Valais	2694Mtp1	WW		
Len de l'El	446Mtp1	WW	B	2008
Lialmigui	2844Mtp1	TE	B	2008
Lisztés feher	1570Mtp1	WE	B	2008
Magdeleine noire des Charentes	0Mtp1195	WW	B	2008
Maingonnat 3 L. 1	2543Mtp1	WE	B	2008
Malvasia istriana	1844Mtp1	WE	B	2008
Massirart	0Mtp698	WW	B	2008
Matrassa blanc faux	0Mtp703	TE	B	2008
Mauzac	443Mtp1	WW	B	2008

Name	Accession number	Sub-population <sup>s</sup>	Cluster group <sup>†</sup>	Year of phenotyping
Mêcle	188Mtp1	WW	B	2008
Médouar*	0Mtp715	TE	B	2008
Mehdik*	2082Mtp1			
Mireille	2713Mtp1	WW	A	2008
Mission	1827Mtp1	WE	A	2008
Molinera gorda	672Mtp1	TE	B	2008
Mondeuse blanche	176Mtp1	WW	B	2008
Mondeuse <sup>§</sup>	161Mtp1		B	2008
Montepulciano	1284Mtp1	WE	B	2008
Morenoa	0Mtp739	WW	B	2008
Morenzi	0mtp1553	WE	C	2008
Morrastel	9Mtp1	WW	B	2008
Moscato di Terracina	577Mtp1	WE	B	2008
Moscato giallo	569Mtp1	WE	C	2008
Mourisco (Collection EVV Amandio Galhano)	0Mtp752	WW		
Mourtaou	2968Mtp1	WW	B	2008
Mourvèdre Goulé	0Mtp406	WE	A	2008
Mskhali	2644Mtp1	TE	B	2008
Muscat à petits grains blancs*	555Mtp2		B	2009
Muscat d'Alexandrie	308Mtp1	TE	C	2008
Muscate (Collection Ravaz)	0Mtp775	WE	B	2008
Négret pounjut	442Mtp1	WW	B	2008
Negru mare	2460Mtp1	TE		
Negru virtos	1667Mtp3	WE	B	2008
Nehelescol	727Mtp1	TE	B	2008
Nero grosso	0Mtp796	TE	B	2008
Nevoeira	0Mtp799	WW	A	2008
Nieddera	2621Mtp1	WE	B	2008
Nincusa	1648Mtp1	WE	C	2008
Noir de Crimée faux	0Mtp811	TE		
Ohanes	1365Mtp1	TE	C	2008
Okatac faux	0Mtp831	WE	A	2008
Öküz gözü faux	0Mtp832	TE		
Olivette Blanche <sup>§</sup>	667Mtp1		D	2009
Name	Accession number	Sub-population <sup>s</sup>	Cluster group <sup>†</sup>	Year of phenotyping
Olivette rose	43Mtp1	TE	B	2008
Opsimo Edessis	2709Mtp1	TE	C	2008
Orbois*	294Mtp1		NA	2009
Öreg Kadarka	0Mtp561	WE	B	2008
Orlovi nokti*	2461Mtp1			
Osteiner	0Mtp835	WW	A	2008
Pagadebiti	114Mtp1	WE	C	2008
Pardina = Pirovano 130	975Mtp1	TE		
Pasiga = CG 26-858	0Mtp1480	TE	B	2008
Peikani	2597Mtp1	TE		
Penouille	349Mtp1	WW	B	2008
Pero Godal	0Mtp875	WW		
Perricone	1314Mtp1	TE	A	2008
Pervenetz praskoveisky*	2651Mtp2		B	2009
Petit Verdot	328Mtp1	WW	A	2008-2009
Piquepoul noir	11Mtp1	WW	B	2008
Plant de Chaufefonds 53 (Fardeau)	0Mtp1512	WW	A	2008
Plant de Pedebnarde 1	0Mtp1293	WW		
Plant de Ponteilla (Jaubert)	0Mtp500	WE		
Plant de Vic 98-N-3 (Collection Torres S.A.)	0Mtp1724	WW		
Plant du Maroc E (Collection Meknès)*	2158Mtp1			
Platyracho	2304Mtp1	WE	A	2008
Pletchistik*	2652Mtp1		NA	2009
Plyto	0Mtp886	TE	B	2008
Posip bijeli	1654Mtp1	WE	B	2008
Poulsard	252Mtp1	WW		
Pozsonyi fehér	2174Mtp1	WE	A	2008
Primitivo	1277Mtp6	WE	B	2008
Raboso piave	1266Mtp2	WW	B	2008
Ramisco	0Mtp952	WW	A	2008-2009
Razachie rosie	433Mtp1	WE	B	2008
Razdani	2112Mtp1	TE		
Ribote rose	0Mtp961	WE	A	2008
Riminèse	124Mtp1	WE	B	2008
Rivairenc = Aspiran noir	1Mtp1	WW	B	2008
Rosa menna di vacca	1662Mtp1	WE	C	2008-2009
Rosaky rose faux	0Mtp982	WE	B	2008-2009
Rossara	1237Mtp1	WE	A	2008
Roublot	227Mtp1	WW	B	2008
Roussaltis <sup>§</sup>	2312Mtp1		A	2008-2009
Roussette basse de Seyssel	0Mtp995	WW	A	2008
Rudezusa	1636Mtp1	WE	B	2008
Ruzevina	1653Mtp1	WE	B	2008
Saint-Laurent	203Mtp1	WW	A	2008
Salicette (Collection Parc de la Tête d'Or)	830Mtp1	WW	NA	2009
San Lorenzo	0Mtp1004	WW	A	2008-2009
Sao Mamede	0Mtp1005	WW	B	2008-2009
Sapéré otskhanouri faux	2655Mtp1	TE		
Sasca	0Mtp1033	WE	B	2008
Savagnin blanc	257Mtp1	WW	A	2008
Semillon	332Mtp1	WW	B	2008
Siah	2074Mtp1	TE	B	2008



Name	Accession number	Sub-population <sup>§</sup>	Cluster group <sup>¶</sup>	Year of phenotyping
Skiadopoulo	0Mtp1068	WE	B	2008
Skylopnichtis	1801Mtp1	WE	NA	2008
Soiaki	2657Mtp1	TE	B	2008
Sorok Let Oktiabria	2683Mtp1	WE	B	2008
Sourkhak biely	2856Mtp1	TE		
Souzao faux	0Mtp1072	WW	A	2008-2009
Staphidampelo	2317Mtp1	WE	B	2008
Starinky	0Mtp1073	WE	B	2008
Stavroto faux (Collection Lykovrissi)	0Mtp1074	WE	C	2008
Strophylatico	2318Mtp1	WE	NA	2008
Sultanina de Bulgarie	2774Mtp1	TE	B	2008
Sultanine*	1566Mtp2			
Syrah*	150Mtp3		B	2009
Tachtas	1815Mtp1	TE	C	2008
Tagobi	2659Mtp1	TE	A	2008
Tantovina	1620Mtp1	WE	B	2008
Tavkveri	1218Mtp1	TE	B	2008
Terret Bouschet	537Mtp1	WW	B	2008
Teta de Vaca	1367Mtp1	TE	C	2008
Tibouren	86Mtp1	WW	B	2008
Totika	0Mtp1129	WE	B	2008
Touriga nacional	1493Mtp1	WW	A	2008
Tsitsa Kaprei*	2471Mtp1	TE	B	2008
Tsolikouri*	0Mtp1150		NA	2009
Tzimliansky belyi	0Mtp1235	WE	A	2009
Ugni blanc	74Mtp1	WE	B	2008
Urbanitraube noir (Collection Oberlin)	0Mtp1154	WE	A	2008
Urmi dinka	0Mtp1156	WE	B	2008
Variété d'oasis Bou Chemma 46*	0mtp1557	TE		
Vassarga bielalaia	126Mtp1	TE	B	2008
Velteliner rouge	284Mtp1	WE	A	2008
Verdea	1258Mtp1	WE	B	2008
Verdeca = Lagorthi	2287Mtp1	WE	B	2008
Verdelho tinto	0Mtp1176	WW	A	2008
Verdelho tinto femelle	1494Mtp1	WW		
Verdiso	1301Mtp1	WW	NA	2009
Verico	1753Mtp1	TE	C	2008
Vermantino nero	2418Mtp1	WE	C	2008
Vidiano	1814Mtp1	WE	B	2008
Vinhão	1483Mtp1	WW	A	2008
Voskeat*	2511Mtp1			
Vulpea faux (Collection Ravaz)	0Mtp1213	WE	A	2008
Wildbacher de Hongrie (Collection Ravaz)	0Mtp1218	WW	A	2008
Zakynthino	3000Mtp1	WE	A	2008
Zeini abiad	1248Mtp2	TE		

The sample of 26 genotypes representative of the 3 sub-populations used for histological observations are in grey boxes.

\* Genotypes maximizing the genetic diversity of cultivated *Vitis vinifera* L. compartment (Le Cunff et al, 2008; Houel et al, 2010)

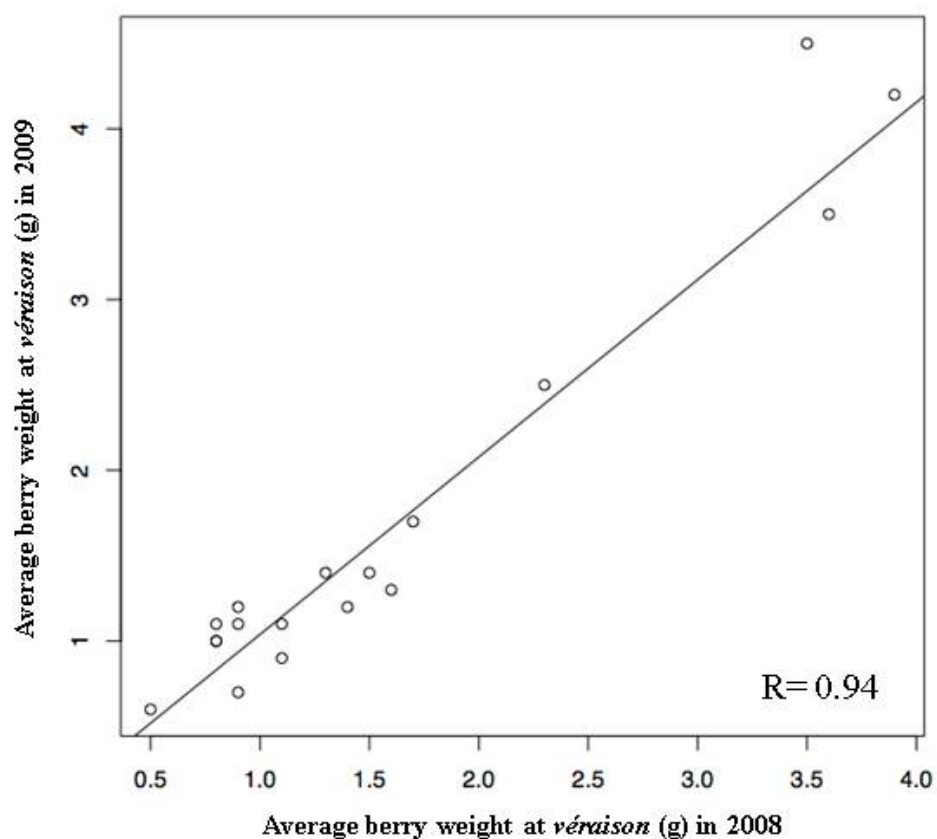
<sup>§</sup> Added genotypes in order to improve the sampling of the phenotypic diversity for berry size in cultivated grapes

<sup>♦</sup> Wild *Vitis vinifera* (Houel et al, 2010)

<sup>§</sup> Structure in the three sub-populations for 278 out of all the genotypes based on 20 SSR by Laucou et al (2011)

(WW: wine grapes from West Europe, WE: wine grapes from East Europe and TE: table grapes from East Europe)

<sup>¶</sup> Genotypes that could not be sorted into a cluster group due to the absence of phenotyping data during 2 years ( ) or due to missing data in one out of the three variables used for clustering (NA).



**Figure S1.** Correlation between the average berry flesh weight at *véraison* in 2008 and 2009 from the 18 genotypes used as control.

Description: Plot of the average berry flesh weight at *véraison* measured in 2008 and the average berry flesh weight at *véraison* measured in 2009 from the 18 cultivars used as control.

**Table S2.** Intra-genotypic average and standard deviation for the parameters measured on three to ten ovaries or berries in 26 genotypes.

Variety name	Accession number	Berry flesh weight (g) at <i>véraison</i> *		Seed number*		Seed weight (g)*		Total mesocarp cell density of ovary (cell.µm <sup>-2</sup> ) <sup>#</sup>	Total mesocarp cell density of berry (cell.µm <sup>-2</sup> ) <sup>#</sup>	Average	Standard deviation
		Average	Standard deviation	Average	Standard deviation	Average	Standard deviation	Average	Standard deviation		
Baresana = Korithi aspro	1247Mtp1	1.35	0.46	2.3	0.9	0.13	0.06	0.0087	0.0019	0.00022	0.00002
Primitivo	1277Mtp6	1.21	0.38	1.7	0.8	0.13	0.06	0.0090	0.0008	0.00036	0.00013
Ohanes	1365Mtp1	0.80	0.23	1.2	0.4	0.10	0.14	0.0123	0.0036	0.00023	0.00000
Alvarelhão	1481Mtp1	2.53	0.44	2.4	0.8	0.15	0.05	0.0093	0.0010	0.00036	0.00015
Carcajolo	1531Mtp1	0.67	0.16	2.2	0.7	0.09	0.03	0.0063	0.0008	0.00033	0.00002
Berbecel	1573Mtp1	0.79	0.21	1.3	0.4	0.06	0.03	0.0090	0.0016	0.00056	0.00018
Kövidinka	1578Mtp1	0.85	0.19	1.4	0.5	0.06	0.03	0.0053	0.0050	0.00060	0.00008
Mondeuse	161Mtp1	2.47	0.44	1.5	0.9	0.11	0.03	0.0101	0.0011	0.00024	0.00008
Chaouch blanc	1673Mtp1	1.40	0.23	2.6	0.7	0.09	0.03	0.0094	0.0007	0.00015	0.00011
Verdeca = Lagorthi	2287Mtp1	1.90	0.38	1.7	0.8	0.09	0.04	0.0081	0.0011	0.00020	0.00003
Roussaïtis	2312Mtp1	1.24	0.20	2.2	0.7	0.09	0.03	0.0097	0.0014	0.00042	0.00018
Alexandroouli	2500Mtp1	0.71	0.13	2.6	0.5	0.07	0.02	0.0104	0.0005	0.00032	0.00011
Savagnin blanc	257Mtp1	3.71	0.95	1.3	0.6	0.10	0.05	0.0070	0.0006	0.00032	0.00002
Fer	353Mtp1	1.26	0.25	2.2	0.7	0.13	0.05	0.0113	0.0008	0.00043	0.00017
Courbu	413Mtp1	0.83	0.17	1.8	0.7	0.08	0.04	0.0107	0.0016	0.00050	0.00014
Razachie rosie	433Mtp1	1.02	0.17	2.7	0.8	0.14	0.05	0.0102	0.0006	0.00032	0.00015
Mauzac	443Mtp1	0.80	0.13	3.0	0.8	0.15	0.03	0.0072	0.0005	0.00034	0.00021
Chasselas	585Mtp1	1.28	0.25	1.5	0.5	0.09	0.03	0.0090	0.0008	0.00040	0.00000
Olivette Blanche	667Mtp1	1.13	0.16	3.0	0.6	0.16	0.04	0.0078	0.0009	0.00014	0.00005
Nehelescol	727Mtp1	1.46	0.31	2.3	1.0	0.11	0.05	0.0096	0.0015	0.00024	0.00007
Dabouki <sup>§</sup>	735Mtp1	–	–	–	–	–	–	0.0117	0.0008	–	–
Heptakilo	743Mtp1	1.43	0.21	2.1	0.5	0.09	0.02	0.0095	0.0007	0.00023	0.00004
Coarna alba	749Mtp1	4.17	0.76	1.8	0.6	0.13	0.06	0.0118	0.0003	0.00036	0.00007
Ugni blanc	74Mtp1	1.36	0.29	1.3	0.4	0.08	0.03	0.0056	0.0010	0.00032	0.00012
Calitor	77Mtp1	1.18	0.22	2.6	0.7	0.20	0.26	0.0091	0.0005	0.00030	0.00016
Ahmar Derani	848Mtp1	3.85	0.66	2.6	1.1	0.18	0.08	0.0095	0.0006	0.00028	0.00018

<sup>§</sup> Missing data due to genotype not producing enough berries

\* Ten berries or ovaries per genotype

<sup>#</sup> Three berries or ovaries per genotype

**Table S3.** Anova test of the difference of berry weight, seed weight, berry cell density and ovary cell density between the four Groups of genotypes, taking into account the intra genotypic variation. Table cells are in grey when the P value is < 0.05 after the Bonferroni correction.

Berry weight

A				
B				
C				
D				
	A	B	C	D

Berry cell density

A				
B				
C				
D				
	A	B	C	D

Seed weight

A				
B				
C				
D				
	A	B	C	D

Ovary cell density

A				
B				
C				
D				
	A	B	C	D

**Annexe 2:** Figures et tableaux supplémentaires de  
l'article: *Screening candidate genes for berry traits using  
association genetics in grapevine (Vitis vinifera L.)* (Chapitre  
4)

**Table S1.** Plant material.

<b>Name</b>	<b>Accession number</b>	<b>Sub-population<sup>§</sup></b>
Achlamiche	2991Mtp1	TE
Aetonychi lefko	2281Mtp1	TE
Ahmar Derani	848Mtp1	TE
Ahmeh Sal apyrène	2982Mtp1	TE
Ak ouzioum tagapskii*	2897Mtp1	—
Alexandroouli	2500Mtp1	TE
Alfrocheiro preto	50Mtp1	WW
Altesse	167Mtp1	WW
Alvarelhão <sup>§</sup>	1481Mtp1	—
Aouillat	403Mtp1	WW
Araklinos*	1805Mtp1	WE
Arakseni tcherni	2503Mtp1	TE
Arbane	219Mtp1	WW
Argant	261Mtp1	WE
Arinarnoa	2747Mtp1	WW
Arinto tinto (Collection Soares Franco)	0Mtp36	WW
Aromriesling	601Mtp1	WW
Arvine	1837Mtp1	WW
Assoued kere	1695Mtp1	TE
Asswad Abou Khislé	2995Mtp1	TE
Assyl kara*	2505Mtp1	TE
Attiki	3005Mtp1	TE
Autumn seedless	2893Mtp1	TE
Avarengo	2218Mtp1	WW
Avgoustiatis	0Mtp64	WE
Azizi el Jaïa	2003Mtp1	TE
B 40-97 (Ramming)	0Mtp1676	TE
Bacchus = Geilweierhof 33-29-133	2725Mtp1	WW
Balbut bijeli	1888Mtp1	WE
Balsamina (Collection Oberlin)	3009Mtp1	WW
Barlinka	2225Mtp5	TE
Baroque	380Mtp1	WW
Baserri	2890Mtp1	WW
Basicata	2298Mtp1	WE
Belle Denise	0Mtp89	WW
Bellone	1338Mtp1	WE
Berbecel <sup>§</sup>	1573Mtp1	—
Beregi rozsas	1629Mtp1	WE
Bianco d'Alessano	2043Mtp1	WE
Blanc Auba	344Mtp1	WW
Blanchier	2886Mtp1	WW
Blank blauer	1631Mtp1	WE
Bonamico	1354Mtp2	WE
Bou Khanzir noir	1909Mtp1	TE
Cabernet franc*	324Mtp1	—
Cacaboué	2070Mtp1	WW
Cahours	0Mtp171	WW

Name	Accession number	Sub-population <sup>s</sup>
Calitor <sup>s</sup>	77Mtp1	—
Camaraou noir	411Mtp1	WW
Carcajolo	1531Mtp1	WW
Carignan	18Mtp1	WE
Catarratto bianco	1303Mtp1	WE
Centennial seedless	2874Mtp1	TE
César*	225Mtp2	—
Chahnani	1744Mtp1	TE
Chami abiad	0Mtp220	TE
Chaouch blanc	1673Mtp1	TE
Chaptal	791Mtp1	TE
Chasselas	585Mtp1	WW
Chatus	129Mtp1	WW
Chirai obak*	1186Mtp1	TE
Chouchillon*	192Mtp1	—
Clairette	26Mtp1	WW
Claverie coulard	0Mtp1656	WW
Coarna alba	749Mtp1	WE
Coda di volpe bianca	1261Mtp1	WE
Codivarta	2972Mtp1	WE
Corason de Kabritte blanc	0Mtp270	WW
Corbeau	157Mtp1	WW
Courbu	413Mtp1	WW
Crimposie	2243Mtp1	WE
Dabouki	735Mtp1	TE
Danugue	666Mtp1	TE
Darkaia noir	629Mtp1	TE
Dimrit noir (Izmir)	2107Mtp1	WE
Domina = Geilweilerhof 4-25-7	0Mtp313	WW
Dureza	139Mtp1	WW
Emerald seedless	2708Mtp1	TE
Espadeiro tinto*	1498Mtp1	—
Fer <sup>s</sup>	353Mtp1	—
Feteasca regala	2247Mtp1	WE
Fokiano	1797Mtp1	WE
Frankenthal rouge foncé	632Mtp1	TE
Freisa	1245Mtp1	WW
Fumin	2371Mtp1	WW
Galbena de Odobesti	2136Mtp1	WE
Galotta	0Mtp1747	WW
Gantziandan	2902Mtp1	TE
Garrido macho	1972Mtp1	TE
Gascon	226Mtp1	WW
Genk Uzum	789Mtp1	WE
Gheméra	329Mtp1	WE
Gibert	372Mtp1	WW
Gouais blanc	210Mtp1	WE
Grain	0Mtp1513	WW

Name	Accession number	Sub-population <sup>s</sup>
Greco bianco	2373Mtp1	WE
Gros Bourgogne	80Mtp1	WE
Gros Cabernet	1232Mtp1	WW
Grosse Méridle	0Mtp537	WW
Hagnos Zöld	0Mtp469	WE
Heptakilo <sup>s</sup>	743Mtp1	—
Heroldrebe = We S 130	2745Mtp1	WE
Heunisch schwarz	1576Mtp1	WW
Hunisa	635Mtp1	TE
Ichkimar	0Mtp586	TE
Impigno	2057Mtp1	WE
INRA Colmar Lignée PN40024	0Mtp1840	reference
Joubertin	154Mtp1	WW
July Muscat	2282Mtp1	TE
Kadarum = Babic (Collection Split)	1644Mtp1	WE
Kaisermuskat	613Mtp1	WW
Kakotryghis	0Mtp562	WE
Kapistroni tétri hermaphrodite (Coll. Kichinev)*	0Mtp567	—
Karitsiotis	1784Mtp1	WE
Kasoufi de la Bekaa	1715Mtp1	TE
Katta-kourgan*	556Mtp1	TE
Khalili belyi	657Mtp1	TE
Khalt	2013Mtp1	TE
Khikhvi	0Mtp581	TE
Kichmich tcherni*	0Mtp583	TE
Kizil sapak faux	746Mtp1	TE
Kolliniatiko	0Mtp603	WE
Kolontar	744Mtp1	WE
Korza erevani	0Mtp610	TE
Kövidinka	1578Mtp1	WE
Koz ouzioum	2635Mtp1	TE
Labrusco	2953Mtp1	WW
Lagrein	1287Mtp1	WW
Lambrusco del Caset	0Mtp633	WW
Lambrusco Marani	2349Mtp1	WW
Lameiro*	0Mtp636	WW
Len de l'El	446Mtp1	WW
Lialmigui	2844Mtp1	TE
Lisztes feher	1570Mtp1	WE
Magdeleine noire des Charentes	0Mtp1195	WW
Maingonnat 3 L 1	2543Mtp1	WE
Malvasia istriana	1844Mtp1	WE
Massirart	0Mtp698	WW
Matrassa blanc faux	0Mtp703	TE
Mauzac	443Mtp1	WW
Mècle	188Mtp1	WW
Médouar*	0Mtp715	TE
Mehdik*	2082Mtp1	—



Name	Accession number	Sub-population <sup>s</sup>
Mireille	2713Mtp1	WW
Mission	1827Mtp1	WE
Molinera gorda	672Mtp1	TE
Mondeuse blanche	176Mtp1	WW
Mondeuse <sup>s</sup>	161Mtp1	—
Montepulciano	1284Mtp1	WE
Morenoa	0Mtp739	WW
Morenzi	0mtp1553	WE
Morrastel	9Mtp1	WW
Moscato di Terracina	577Mtp1	WE
Moscato giallo	569Mtp1	WE
Mourtaou	2968Mtp1	WW
Mourvèdre Goulé	0Mtp406	WE
Muscat à petits grains blancs*	555Mtp2	—
Muscat d'Alexandrie	308Mtp1	TE
Muscate (Collection Ravaz)	0Mtp775	WE
Négret pounjut	442Mtp1	WW
Negru virtos	1667Mtp3	WE
Nehelescol	727Mtp1	TE
Nero grosso	0Mtp796	TE
Nevoeira	0Mtp799	WW
Nieddera	2621Mtp1	WE
Nincusa	1648Mtp1	WE
Ohanes	1365Mtp1	TE
Okatac faux	0Mtp831	WE
Olivette Blanche <sup>s</sup>	667Mtp1	—
Olivette rose	43Mtp1	TE
Opsimo Edessis	2709Mtp1	TE
Orbois*	294Mtp1	—
Öreg Kadarka	0Mtp561	WE
Orlovi nokti*	2461Mtp1	—
Osteiner	0Mtp835	WW
Pagadebiti	114Mtp1	WE
Pasiga = CG 26-858	0Mtp1480	TE
Penouille	349Mtp1	WW
Perricone	1314Mtp1	TE
Pervenetz praskoveïsky*	2651Mtp2	—
Petit Verdot	328Mtp1	WW
Piquepoul noir	11Mtp1	WW
Plant de Chaudefonds 53 (Fardeau)	0Mtp1512	WW
Plant du Maroc E (Collection Meknès)*	2158Mtp1	—
Platyracho	2304Mtp1	WE
Pletchistik*	2652Mtp1	—
Plyto	0Mtp886	TE
Posip bijeli	1654Mtp1	WE
Pozsonyi fehér	2174Mtp1	WE
Primitivo	1277Mtp6	WE
Raboso piave	1266Mtp2	WW

Name	Accession number	Sub-population <sup>s</sup>
Ramisco	0Mtp952	WW
Razachie rosie	433Mtp1	WE
Ribote rose	0Mtp961	WE
Riminèse	124Mtp1	WE
Rivairenc = Aspiran noir	1Mtp1	WW
Rosa menna di vacca	1662Mtp1	WE
Rosaky rose faux	0Mtp982	WE
Rossara	1237Mtp1	WE
Roublot	227Mtp1	WW
Roussaïtis <sup>s</sup>	2312Mtp1	—
Roussette basse de Seyssel	0Mtp995	WW
Rudezusa	1636Mtp1	WE
Ruzevina	1653Mtp1	WE
Saint-Laurent	203Mtp1	WW
Salicette (Collection Parc de la Tête d'Or)	830Mtp1	WW
San Lorenzo	0Mtp1004	WW
Sao Mamede	0Mtp1005	WW
Sasca	0Mtp1033	WE
Savagnin blanc	257Mtp1	WW
Semillon	332Mtp1	WW
Siah	2074Mtp1	TE
Skiadopoulo	0Mtp1068	WE
Soïaki	2657Mtp1	TE
Sorok Let Oktiabria	2683Mtp1	WE
Souzao faux	0Mtp1072	WW
Staphidampelo	2317Mtp1	WE
Starinky	0Mtp1073	WE
Stavroto faux (Collection Lykovrissi)	0Mtp1074	WE
Strophyliatico	2318Mtp1	WE
Sultanina de Bulgarie	2774Mtp1	TE
Sultanine*	1566Mtp2	—
Syrah*	150Mtp3	—
Tachtas	1815Mtp1	TE
Tagobi	2659Mtp1	TE
Tantovina	1620Mtp1	WE
Tavkveri	1218Mtp1	TE
Terret Bouschet	537Mtp1	WW
Tibouren	86Mtp1	WW
Totika	0Mtp1129	WE
Touriga nacional	1493Mtp1	WW
Tsitsa Kaprei*	2471Mtp1	TE
Tsolikouri*	0Mtp1150	—
Tsolikouri*	0Mtp1150	—
Tzimliansky belyi	0Mtp1235	WE
Ugni blanc	74Mtp1	WE
Ugni Blanc mutant	0Mtp1596	—
Urbanitraube noir (Collection Oberlin)	0Mtp1154	WE
Urmi dinka	0Mtp1156	WE

<b>Name</b>	<b>Accession number</b>	<b>Sub-population<sup>§</sup></b>
Vassarga bielaia	126Mtp1	TE
Velteliner rouge	284Mtp1	WE
Verdea	1258Mtp1	WE
Verdeca = Lagorthi	2287Mtp1	WE
Verdelho tinto	0Mtp1176	WW
Verdiso	1301Mtp1	WW
Verico	1753Mtp1	TE
Vidiano	1814Mtp1	WE
Vinhão	1483Mtp1	WW
Voskeat*	2511Mtp1	
Wildbacher de Hongrie (Collection Ravaz)	0Mtp1218	WW
Zakynthino	3000Mtp1	WE

Genotypes used for quantitative Real-Time PCR assays are in grey box

\* Genotypes of the SNP discovery pannel

§ Structure in three sub-populations (WW: wine grapes from West Europe, WE: wine grapes from from East Europe and TE: table grapes from East Europe, ( ) genotypes which do not belong to the 3 sub-populations (Laucou et al, 2011; R. Bacilieri personal communication)

§ Added genotypes in order to improve the sampling the phenotypic diversity for berry size

**Table S2.** List of the candidate genes sequenced for SNP discovery and primers used.

Gene id 12X Genome Assembly	Gene name	Chromosome	Fragment name	Primer forward	Primer reverse	Fragment size	Sequencing*
GSVIVG01008140001	<i>VvAGL8</i>	17	VVC0011A	CCAAGATTATTAACACAAAGCA	TCCCTACCTTTTTCATATCC	408	F
			VVC0011B	CTTAGCTTGGTCTTACCTTTTT	ACTGCATGTTGTGCTTCA	704	F
			VVC0011C	AAACCTTAATTTTGGGAATG	TCCTAATTGATTTGATGTTGAG	811	F/R
GSVIVG01013435001	<i>VvHMGR1</i>	18	VVC2942A	GATCGCATCCTTCATCTATC	TGAATACTTTGCAGCCTTG	1296	F/R
			VVC2942F	AAATGTTTCTTCATCCCTTC	CACTGCCCTAAAATTGCTT	804	R
			VVC2942G	GTTCAGACAAGAAACCAGCT	TGACAGAAAGCAATTCACA	1032	F/R
			VVC2942H	GATTAGCAGCCAGAAAGTG	GATCCTTCCCATCATTAACA	1398	F/R
			VVC2942I	GGTAGCCAGTACCAACAGA	TGAATACTTTGCAGCCTTG	765	F/R
			VVC2942S	ATCGTCCCGGAAACTACTAA	GTTTCTTCATCCCTTCGAG	656	F/R
			VVC6002A	GTAGTCGAACCCATCCAAG	ATATCTCCGCTTCTCCTTTC	735	F/R
GSVIVG01026444001	<i>VvHMGR2</i>	4	VVC6002B	CATGTTAAGCTCCACCAGAG	AGAGGAGGAAGAGGAGATGA	1146	F/R
			VVC6002E	GGTGATGCAATGGGTATG	ACATCTCTGCTGGATCTGTT	860	F/R
			VVC6002D	AGTCCCTATATCTCCGCTTC	CGTCACAGGAGCAACATC	562	R
			VVC7282A	TCCAACCTCATCTCTCCATTC	TCACAACCTTCTCAACATCA	272	F/R
–	<i>VvPMEI2</i>	2	VVC7282B	TTTCCATTCTTGTGTGTCAA	TTAGCCTGTGTGGTGTATG	724	F/R
GSVIVG01009815001	<i>VvAP3</i>	18	VVC14506A	TTTCGTATCCTTCTCTCACC	GAAGTTGGAGTTGTTCTGCT	1217	F/R
			VVC14506C	AATTTTGAGTGAGCATTTT	AAAACCTCAACAGGTCCATTAG	659	F/R
GSVIVG01009691001	<i>VvOvate2</i>	18	VVC14634A	GAAGGTGTTGAAAGTGGAAG	GCGTGTGGATCATATCTCTAA	1304	F
			VVC14634B	ATCGGGTGTAAGGTCTGAA	AGAATGGTTTGAAGTTGCTG	929	F
GSVIVG01009219001	<i>VvAGL16</i>	18	VVC15065D	GCTATTGCAGTTGTAGTCTGCT	CGTTGGAAGAAATGGTCTC	925	F
			VVC15065E	AGAAAACCTAGCTTGGAGGT	CAAAAAGAAGCATAATAGAGATTGA	331	F
			VVC15065F	TTGAAAATTTTATCCATGTATGTTC	CATGCAAAATAGCAAGTGG	263	R
			VVC15065G	GAAATGAGTCTTCGAGGTG	AAATACTTGGGGTATCTTTGA	678	F/R
GSVIVG01009171001	<i>VvAGL22</i>	18	VVC15108A	GCTTCGTTTATTGCTTTGA	TTTGGGCATTCTATGTGAGT	404	F
			VVC15108B	CCGACTAAGATGAGGTGAAG	AGCTGTTTGATATGATTTTGA	928	F/R
GSVIVG01008806001	<i>VvMADS9</i>	18	VVC15451C	AGACAAATATAGGCCAACAAG	TCAACTCTTCTGATCTCTTCA	547	F/R
			VVC15451B	GAGTTATATCCTCTCCTGTAAGTTG	TTTTCAATTTGTTTCATGTTCC	548	F/R
			VVC15452A	GACAGGCAAGAACTCCA	GGAAGAGGAGGAAGAAGAGA	527	F
GSVIVG01037533001	<i>VvYABBY2</i>	6	VVC17462A	ACAATCACCACTGTTTTCC	AACCATTGCTACCGTGATT	772	F
			VVC17462G	CTAAATTGAAAGCCCTTGGT	CATCTTCTCTGGCTCAACTC	329	F
			VVC17462C	TTTTTCTCCTGGTCATCATC	GTGACGTGTAATTAATTTGTTG	338	F
			VVC17462F	GAGATCCAAGTGAGGATTGT	TTGTATAGCCCAAAATATCAAG	1316	R
GSVIVG01033891001	<i>VvSPY</i>	8	VVC21566H	AGCCTTTTGTGTGTAGTGTG	GATGCACCTACTAAGATTGGA	1634	F/R
			VVC21566I	TGGAATTGACGAGAAGAAG	TTAATGGCAAAAGATCAACA	1362	F/R
			VVC21566A	ATGTCGGATAATGAGGGTTT	TGGCAGAGTAACAAACAAGA	651	F
			VVC21566B	CAACCCAAAGAAAGAAAGAA	GTTTCAGGTGCTAACAAAAAGA	775	F/R
			VVC21566C	TATAGGTTAAATTGGAAGGAGAC	AACCAACAATATCAATAGAAAGA	753	F/R
			VVC21566D	TGCTTTTTGTATCCTTGTTTTT	TGCCAAGTAGATATGATTGAAC	1667	F/R
			VVC21566E	CCACCTTGGAAGTGCTTAGT	AGCCAGGATAAAAAGAAAAGA	2008	F/R
			VVC21566S	CACGCAATTCATTATAGGA	CTTCGATGTTTTATTTTGTTC	468	F
GSVIVG01033743001	<i>VvRPL24</i>	8	VVC21714C	TGTGTTTTGGTCACTTTCTG	CAACCTACCGATCTCACTCT	656	F/R
			VVC21714A	CTTCCTTTTCTAGGGTTTCC	AATTTCTACAGCACCCAAAA	1495	F
			VVC21714D	GCTTCCTTTTCTAGGGTTTC	TCCTTTCCAGGGTATATCT	1034	F

Gene id 12X Genome Assembly	Gene name	Chromosome	Fragment name	Primer forward	Primer reverse	Fragment size	Sequencing*
GSVIVG01025916001	<i>VvAGL12</i>	18	VVC21903A	ACGTTAAGAGGCAACAGAGA	GTAAAACGACCTCCCATACA	789	F/R
			VVC21903B	CGTTTCTTAAAATCAGTATCAAAG	GGTATTCATAACCCTCCAGTT	1280	F
			VVC21903C	CAACCACCACCCAGTAGTAA	TTAGTTAGCTTCCAAGAGCA	405	R
			VVC21903D	AAACTCCTATACAAAATCCATTG	TTCATATATTGGAATGCAAAAC	410	F
GSVIVG01025945001	<i>VvAGL11</i>	18	VVC21934A	TTGGTGCAATAAACCTAACA	GTTGAAGAAATGGTGGAGAA	242	F/R
			VVC21934B	CTCCTTTAGGGTTTTGCTTT	CGGAGATATACGCTTTCATT	1337	F/R
			VVC21934C	TCCATATCCATACTACTCTTCA	AGGCTACTTGGTGATTATGT	301	F
			VVC21934D	CTAAAGCCTGATGGTAGGAA	AAAGACGAAGAAATCTCACAA	897	F
			VVC21934E	CTACCAGCAAGAATCAGCA	GTACTCAATCTCAGCCAACAG	703	F/R
GSVIVG01018989001	<i>VvOvate1</i>	4	VVC24966A	TTCTTTTTCTTACGCCTCAC	TTGAAGGCACTCCTCACT	804	F/R
			VVC24966B	ATGTTCTTCAGAGAGCTTTAGAC	CAGGAGAATGACGGTGAG	559	R
			VVC24966C	TTGAAGGCACTCCTCACT	GGTTGTCGTCGTAGACCTT	485	R
GSVIVG01033546001	<i>VvSUN</i>	8	VVC25335A	TCTTCATCCGCTCTCTTT	AAACCAAGGTGTGCTTGA	1336	F/R
			VVC25335B	GAAGTATGTCGGGAACTAA	ATTGGGAGGGAGAGAAATAA	1077	F
			VVC25335C	TCAATAACCAAGTGGAGTGTA	GAAGTATGTCGGGAACTAA	642	F/R
GSVIVG01023852001	<i>VvHMGR3</i>	3	VVC36301B	GAAGGAAATCCAAAACATTC	AAGACGAGGAAATCATCAAG	737	F/R
			VVC36301C	GGTTTATAGTGGCCTCAAGAT	TGCAGGGAAGAACCTGTA	956	F/R
			VVC36301F	AAAGCCATCTACGCCTCT	AGTGAGGTTCTTGAGCATGT	709	F/R
			VVC36301G	CATCTGCAGTCATGGATGT	CCCAAAATCGACTCGTAA	673	F/R
GSVIVG01031755001	<i>VvFw2.2.3</i>	3	VVC36579A	ACAAGTTGCAGTAGCACACA	CAGATTGGGTGAAGGCTAT	860	F/R
			VVC36579B	CCAAAGGATTGTCAAAGG	TAAGGAGCAACTAGGCAAGA	425	F
GSVIVG01031754001	<i>VvFw2.2.1</i>	3	VVC36581A	ACCTATTCCTCATGTCAAATC	GGAGCTACCAAGCAAGAC	730	F/R
			VVC36581C	GGCCTGCTTGATTCTTGT	CTCCATAGAGTTGAAGATCCA	997	F/R
GSVIVG01031753001	<i>VvFw2.2.2</i>	3	VVC36582A	GGTGACCAATTCATTTCACT	CATCCACTGATCCACAAATC	1544	F/R
			VVC36582C	GTGAGTGGTACGATCTATACGG	AAAAACCCCATCTTCAACT	294	F/R
			VVC36582B	GCAACTAGGAAAGACTTGGA	ACATCCACTGATCCACAAAT	1575	F
			VVC36582S	ACATATCACCAATAAACTGGA	AATGCTATAACTTGTGGATGG	346	F
GSVIVG01020033001	<i>VvHB13</i>	1	VV05013A	GAGGATGATTGTCTGATG	GCTTCTCTTTGACTGG	1162	F
			VVC37014A	GAACCAGTTAAGGACATAGGG	TGGGTCTCTCAATTCATTTT	842	F
			VVC37014B	TTCTCTATTGCGTAGTTGG	ATCATTTTGGATGGGATTAG	512	F/R
GSVIVG01019761001	<i>VvMJ</i>	2	VVC0413D	GCATAACTGGTGATTAAGGTTT	GAGCTGAAACCATAAGAATCA	1523	F
GSVIVG01010521001	<i>VvMADS4</i>	1	VVC2777C	TGCACTCAACAAAATCTCC	GGCCTTCCTCTTATTATGGT	586	F
GSVIVG01000802001	<i>VvMADS1</i>	12	VVC18932C	ATGATGATTTCCTAAGACTGTG	CAATTCGAGCCAATATTC	631	F
GSVIVG01031404001	<i>VvHD2C</i>	6	VVC22095A	CAAAGGTGAAGGTTGAAGAG	GGCCTTGCATCAATAACTAA	1070	F
GSVIVG01031943001	<i>VvSVP2</i>	3	VVC36409A	TTAATTGATTGAGAACATGGAG	TGAAAGGCATACATAGACCA	630	F
GSVIVG01031941001	<i>VvBURP1</i>	3	VVC36411A	GACATTTCTCCCCATTTTT	GCAACATTCAAAACCTACTCA	1084	F
GSVIVG01036549001	<i>VvMADS6</i>	14	VVC38079A	CTTCTGCCTTCTTTCTTCA	CAAACCAAAACCAAAACAC	397	F

\* Forward (F) and/or reverse (R) sequencing orientation

**Table S3.** List of the SNP in the candidate genes for berry size variation used for genotyping.

Gene id	Gene name	Chromosome	Fragment name	SNP type* <sup>s</sup>	SNP position	SNP position on the chromosome
GSVIVG01008140001	<i>VvAGL8</i>	17	VVC0011B	[G/A]	Intron	5435479
			VVC0011C	[T/C]	Exon	5434897
				[G/A]	Intron	5434496
GSVIVG01013435001	<i>VvHMGR1</i>	18	VVC2942A	[T/G]	Intron	497611
				[G/C]	Exon	497550
				[T/C]	Exon	497314
			VVC2942F	[T/G]	Intron	497611
				[G/T]	Intron	498038
			VVC2942G	[C/T]	Exon	495197
				[G/A]	Intron	495082
			VVC2942H	[C/T]	Intron	496008
GSVIVG01026444001	<i>VvHMGR2</i>	4	VVC6002A	[C/T]	Exon	23597983
				[G/C] <sup>s</sup>	Exon	23597839
			VVC6002B	[C/T]	Exon	23598326
—	<i>VvPMEI2</i>	2	VVC7282B	[G/C]	—	23342460
				[G/C]	—	23342683
				[G/C]*	—	23342463
GSVIVG01009815001	<i>VvAP3</i>	18	VVC14506A	[T/C]	Exon	11506795
				[G/T]	Intron	11507484
			VVC14506C	[T/C]	Intron	11511798
				[C/T]*	Exon	11511848
GSVIVG01009691001	<i>VvOvate2</i>	18	VVC14634A	[C/A]*	Exon	10517794
			VVC14634B	[G/A]	Exon	10517387
				[A/C]	Exon	11305484
				[T/C]	Exon	10517326
GSVIVG01009219001	<i>VvAGL16</i>	18	VVC15065F	[A/C]	Intron	6370312
				[C/T]	Intron	6370342
			VVC15065G	[G/A]	Intron	6375513
				[A/G]	Exon	6375928
GSVIVG01009171001	<i>VvAGL22</i>	18	VVC15108B	[C/T]	Intron	5709089
				[A/G]*	Intron	5709147
				[C/T]	Intron	5709249
				[A/G]	Intron	5709281
				[A/T]	Intron	5709549
				[A/G]	Exon	2293044
GSVIVG01008806001	<i>VvMADS9</i>	18	VVC15451C	[A/G]	Exon	2293044
			VVC15451B	[T/C]	Intron	2292167
				[C/T]	Intron	2292167
			VVC15452A	[A/C]	Exon	2293629
				[A/C]	Exon	2293629
GSVIVG01037533001	<i>VvYABBY2</i>	6	VVC17462A	[-/AATTT]*	Intron	11953941
			VVC17462C	[G/A]	Intron	11954050
				[C/A]	Exon	11954227
			VVC17462F	[T/C]	Intron	11955029
				[C/T]	Intron	11955105
GSVIVG01033891001	<i>VvSPY</i>	8	VVC21566H	[A/G]*	Intron	16966466
			VVC21566I	[A/G]*	Intron	16966822
			VVC21566A	[G/A]	Exon	16982954
			VVC21566B	[A/G]	Exon	16981487
			VVC21566C	[G/A]	Exon	16978735
			VVC21566D	[A/G]	Intron	16976820
			VVC21566E	[G/A]	Exon	16969585
				[G/A]	Exon	16969569
GSVIVG01033743001	<i>VvRPL24</i>	8	VVC21714A	[C/A]	Intron	18194567
				[A/T]	Intron	18194874
				[A/G]	Intron	18194912
			VVC21714D	[T/A]	Intron	18194341
				[C/T]*	Intron	18195610
				[T/G]	Intron	18195462

Gene id	Gene name	Chromosome	Fragment name	SNP type* <sup>s</sup>	SNP position	SNP position on the chromosome
GSVIVG01025916001	<i>VvAGL12</i>	18	VVC21903A	[G/A]	Exon	27268985
				[G/C]	Exon	27269351
				[T/C]	Intron	27269431
			VVC21903B	[G/A]	Intron	27282604
				[G/A]	Intron	27282742
				[C/T]	Exon	27283561
			VVC21903D	[A/G]	Intron	27283336
				VVC21934B	[G/A]	Exon
GSVIVG01025945001	<i>VvAGL11</i>	18	VVC21934D	[G/T]	Intron	26889822
				[C/T]	Intron	26889528
			[C/T]	Intron	26889919	
			VVC21934F	[T/C]	Intron	26890375
			GSVIVG01018989001	<i>VvOvate1</i>	4	VVC24966C
GSVIVG01033546001	<i>VvSUN</i>	8	VVC25335A	[T/C]	Exon	19852424
			VVC25335C	[G/A]	Exon	19853020
				[A/C]	Exon	19853039
				VVC36301B	[A/G]	Exon
GSVIVG01023852001	<i>VvHMGR3</i>	3	VVC36301G	[C/T]	Exon	2970664
				[T/C]*	Exon	2970831
			[G/A]	Exon	2971051	
			VVC36581A	[A/T]	Intron	4068252
			GSVIVG01031754001	<i>VvFw2.2.1</i>	3	VVC36581C
[C/T]	Intron	4067318				
[T/G]	Exon	4067284				
GSVIVG01020033001	<i>VvHB13</i>	1	VVC37014A	[A/C]	Exon	11306700
GSVIVG01019761001	<i>VvJMJ</i>	2	VVC0413D	[T/A]	Exon	3117781
				[C/T]	Exon	3117982
GSVIVG01010521001	<i>VvMADS4</i>	1	VVC2777C	[T/*]	Intron	21369619
				[T/C]	Exon	21369421
GSVIVG01000802001	<i>VvMADS1</i>	12	VVC18932C	[T/C] <sup>s</sup>	Exon	282565
				[T/C]	Exon	282397
GSVIVG01031404001	<i>VvHD2C</i>	6	VVC22095A	[G/C]	Intron	19044289
				[A/G]	Intron	19043977
GSVIVG01031943001	<i>VvSPV2</i>	3	VVC36409A	[C/T]*	UTR	5591281
				[C/T]*	Intron	5591057
GSVIVG01031941001	<i>VvBURP1</i>	3	VVC36411A	[G/T]	Exon	5580220
				[G/A]	Exon	5580132
GSVIVG01036549001	<i>VvMADS6</i>	14	VVC38079A	[C/G]	Exon	23340943

\*The genotyping results were discarded because the primers targetted several loci.

<sup>§</sup> The genotyping results were discarded because the water control was contaminated.

**Table S4.** List of the SNP chosen in 48 random gene fragments.

Gene id	Chromosome	Fragment name	SNP type	SNP position	SNP localisation on chromosome
GSVIVG01012159001	1	VV00984A	[T/C]	Exon	963439
GSVIVG01011586001	1	VV01103A	[T/C]	Exon	6228864
GSVIVG01010320001	1	VV04811A	[A/G]	Exon	19474111
GSVIVG01013247001	2	VV01845A	[T/G]	Intron	6290922
GSVIVG01013145001	2	VV05338D	[A/T]	Intron	7634019
GSVIVG01019858001	2	VV05469A	[A/G]	Exon	3895015
GSVIVG01031752001	3	VV01652A	[G/C]	Intron	4058870
GSVIVG01035578001	4	VV01212A	[T/C]	Exon	2481774
GSVIVG01031674001	5	VV04641A	[T/C]	Exon	19978989
GSVIVG01029057001	5	VV06623D	[T/G]	Exon	11590035
GSVIVG01024773001	6	VV00594A	[C/T]	Intron	7073698
GSVIVG01037436001	6	VV02212A	[C/T]	Exon	13774138
GSVIVG01031397001	6	VV03090A	[G/T]	Intron	19129522
GSVIVG01011061001	7	VV03374A	[A/T]	Exon	1888687
GSVIVG01025719001	8	VV03305A	[G/A]	Exon	12680433
GSVIVG01033277001	8	VV03568A	[A/T]	Exon	21938051
GSVIVG01033690001	8	VV05408A	[C/T]	Exon	18638946
GSVIVG01016126001	9	VV03486A	[C/G]	Exon	19505781
GSVIVG01029529001	9	VV04275A	[G/A]	Exon	21918311
GSVIVG01021562001	10	VV03977A	[C/T]	Exon	7096663
GSVIVG01021753001	10	VV04308A	[A/T]	Exon	10553708
GSVIVG01032233001	11	VV01227D	[C/T]	Intron	13510497
GSVIVG01015569001	11	VV02145A	[A/G]	Intron	5035997
GSVIVG01030508001	12	VV00047B	[C/T]	Exon	6311270
GSVIVG01022919001	12	VV03758A	[G/A]	Exon	18056690
GSVIVG01029753001	12	VV04431A	[A/T]	Intron	14187728
GSVIVG01016585001	13	VV02484A	[G/A]	Intron	2616281
GSVIVG01032161001	13	VV04061A*	[G/T]	UTR	21985886
GSVIVG01034690001	13	VV04784A	[T/C]	Intron	7789750
GSVIVG01032903001	14	VV01715A	[A/C]	Exon	23968999
GSVIVG01032435001	14	VV02778A	[T/G]	Intron	27363346
GSVIVG01021832001	14	VV04364A*	[T/A]	Exon	6758917
GSVIVG01027057001	15	JAR01-05	[A/G]	Exon	18285492
GSVIVG01018223001	15	VV02609A	[G/A]	Intron	13020721
GSVIVG01019302001	15	VV03248A	[T/G]	Intron	1644046
GSVIVG01013967001	16	VV04325A	[G/A]	Exon	8133667
GSVIVG01038685001	16	VV05031A	[C/T]	UTR	21021351
GSVIVG01010605001	16	VV06195A	[T/C]	Exon	16096256
GSVIVG01008052001	17	SID02-14-16	[C/T]	Intron	6220281
GSVIVG01008217001	17	VV02267A	[A/G]	Exon	4545903
GSVIVG01008599001	17	VV02368A	[C/T]	Intron	516821
GSVIVG01010035001	18	VV00006A	[A/T]	Intron	13683687
GSVIVG01012775001	18	VV04618D	[A/G]	Exon	23444276
GSVIVG01037687001	19	VV00797A	[A/T]	Exon	6893676
GSVIVG01000622001	19	VV04667D	[A/G]	Exon	14014476
GSVIVG01014835001	19	VV05852A	[G/A]	Exon	10336709
GSVIVG01007007001	Random	VV06625A*	[G/A]	Intron	29701262
GSVIVG01003182001	Random	VV06645A	[C/G]	Exon	7646974

\*The genotyping results were discarded because the primers targetted several loci.



**Annexe 3:** Figures et tableaux supplémentaires de  
l'article: *Patterns of sequence polymorphism in the Fleshless  
berry locus in cultivated and wild Vitis vinifera  
accessions* (Chapitre 5)

**Additional file 1: supplemental Table S1.** Plant material.

Name	Accession number	Wine (W)/Table (T) /wild (S)	Average berry weight (g) <sup>s</sup>
Ak ouzioum tagapskii*	2897Mtp1	T	3.8
Araklinos*	1805Mtp1	W	2.1
Assyl kara*	2505Mtp1	T	3.0
Cabernet franc*	324Mtp1	W	1.7
César*	225Mtp2	W	2.3
Chirai obak*	1186Mtp1	T	3.6
Chouchillon*	192Mtp1	W	1.8
Espadeiro tinto*	1498Mtp1	W	1.9
INRA Colmar Lignée PN40024	0Mtp1840		
Kapistroni tétri hermaphrodite (Coll. Kichinev)*	0Mtp567	T	3.5
Lameiro*	0Mtp636	W	1.1
Katta-kourgan*	556Mtp1	T	9.2
Plant du Maroc E (Collection Meknès)*	2158Mtp1	T	5.9
Médouar*	0Mtp715	T	5.1
Mehdik*	2082Mtp1	T	2.6
Muscat à petits grains blancs*	555Mtp2	W	2.3
Orbois*	294Mtp1	W	1.7
Orlovi nokti*	2461Mtp1	T	4.3
Pervenetz praskoveïsky*	2651Mtp2	T	2.1
Kichmich tcherni*	0Mtp583	T	2.5
Pletchistik*	2652Mtp1	W	2.0
Sultanine*	1566Mtp2	T	1.9
Syrah*	150Mtp3	W	1.5
Tsitsa Kaprei*	2471Mtp1	T	4.1
Tsolikouri*	0Mtp1150	W	2.4
Variété d'oasis Bou Chemma 46*	0mtp1557	T	?
Voskeat*	2511Mtp1	W	?
Lambrusque Campmarcel 2	8500Mtp23	S	1.0
Lambrusque Colmar femelle	8500Mtp9	S	1.0
Lambrusque Dirmstein mâle	8500Mtp38	S	
Lambrusque Grésigne 1	8500Mtp124	S	1.0
Lambrusque Pic Saint-Loup 2	8500Mtp113	S	1.0
Lambrusque Sejnene 1	8500Mtp31	S	1.0
Lambrusque Teulere	0Mtp1323	S	1.0
Chardonnay	199Mtp12	W	1.7
Pinot Noir clone ENTAV-INRA 777	ENTAV1617	W	1.4
Ugni Blanc	74mtp31	W	2.4
Ugni Blanc mutant (fleshless berry)	0Mtp1596		0.3

<sup>s</sup> The average berry weights at maturity were retrieved from unpublished experiments of C. Houel, L. Torregrosa and from the data archives of the Domain of Vassal

\* genotypes maximizing the genetic diversity of the cultivated *Vitis vinifera* compartment, from Le Cunff et al [23]

**Additional file 2: supplemental Table S2.** Markers and sequence fragments along the *F1b* region. List and localisation on the grapevine genome sequence of the fragments sequenced (SEQ) and of the markers used for genotyping (SSR, CAPS, SNP) together with the primers used for amplification.

12X Genome Assembly			Fragment id	Marker type	Primer pairs <sup>§</sup>		Set of genotypes sequenced or genotyped	
Scaffold n°	Gene id	Position			Forward	Reverse	SnpGrapMap*	CHA-UB*
122	GSVIVG01013479001	27097..27590	VVC2985A	SEQ	AACGAAGGGAGGAAAACC	CAACTCCATATCCAATTTTGTC	yes	
122	GSVIVG01013476001	85127..85871	VVC2982A	SEQ	GCTAAAGAGCTTGGGAAGA	GTTTGAAGATTTTGGTGAGG	yes	yes +R
122	GSVIVG01013475001	94259..94960	VVC2981A	SEQ	TACACAAACACAGACCCAGA	TGGTGGAGGAAATAGAGATG	yes	
122	GSVIVG01013472001	140495..141081	VV05807A	SEQ	AACAGAACGAACAGGAAC	CTGAGATAGGCTTGATTG	yes	
122	GSVIVG01013471001	161551..161929	VV05806A	SEQ	CCTAAGTCTGAACATAAGAAGA	CTGAAACACCTCCAATAAC	yes	yes +R
122	GSVIVG01013469001	176785..176983	VVC2975A	SEQ	TCAAACAAACTTGCTGAATG	AGCTTGATAAACCCAGATGA	yes	
122	GSVIVG01013468001	211001..211674	VVC2974A	SEQ	GGAAGTAAGAAGCGTGAGAA	GTTATCAATATCCCCACAA	yes	yes
122	GSVIVG01013467001	212555..212699	VVS50	SSR	ACCAAATGTTGTTTCTGATTG	TGAATTGGAGACAAAAATAAGC		yes
122	GSVIVG01013466001	213864..214083	VVS51	SSR	TCTGCATTGTTGCAGATGCT	TGTTTCGGACCACCTTTTCA		yes
122	GSVIVG01013462001	226346..226575	VVS52	SSR	ATGTGCTCTTGGAACTCACC	TTTTCAAGGGAACATTTATAGC		yes
122	GSVIVG01013467001	227461..228070	VVC2973A	SEQ	GCGTGTAGGCAGAAAAGA	AATTACGAGATGGTGGATG	yes	
122	GSVIVG01013466001	230668..230769	VVS53	SSR	TTGGGTGAATTTTAAGAAATGG	TCAGGGCCCTACAGGATTAG		yes
122	GSVIVG01013466001	232290..232641	VVC2972A	SEQ	TGAATGAAAGAAGATGTGTTG	CTGGAACATTTTGCTGAGG	yes	
122	GSVIVG01013462001	250896..251595	VVC2970A	SEQ	ATTTTTATTTGGTTGGGTGA	GATTATGGTAAGGGCTGATG	yes	
122	GSVIVG01013461001	261445..262084	VV05805A	SEQ	CCTATCTCTTCAACTCCA	GGTAATCAAGCCATTCTC	yes	yes +R
122	GSVIVG01013460001	284473..284921	VV05804A	SEQ	AGTCTGCTGTGTCTTCTT	TAGTTCCTTCGGCATTTT	yes	
122	GSVIVG01013458001	299201..299664	VVC2967B	SEQ	CAGGAAATAAGTCCAGGATG	ATGGATCACATAACGAAAGG		yes +R
122	GSVIVG01013458001	307808..308363	VVC2966A	SEQ	TATTATGGACACACCCACCT	GGGTTAATCATGTGAAACGAC	yes	
122	GSVIVG01013458001	308176..308256	1036L11F	SNP	GCAGTAGCGGTGTGCTGCGTGTGTCATCTTGTAGGCAGAAGCTGACTCGCCCC TTCCGGT[A/T]ACCAAAGAAGGTGCCAGTCACAGTGGAGGCGTGGGTGGACTC AGTATCACTGGAGATGGA			yes
122	GSVIVG01013456001	314510..315008	VVC2965A	SEQ	GGGTGGATGTGTATCTTCTG	GGAGCCTTCGTAAATGGT	yes	
122	GSVIVG01013454001	321067..321135	VMC3E5	SSR	GATTGTCTTTACAAGGCGTTC	CCAGGAGACTTGCTTTGTATTT		yes
122	GSVIVG01013455001	321452..321822	VV05803A	SEQ	CCCTCTGTTCCTACTGTT	GAATACCTTTTCTTCTCC	yes	yes +R
122	GSVIVG01013454001	325978..326580	VVC2963A	SEQ	ACCAGGGTCCTACAATGAGT	GTAAGGGAGCTTATCGAACC	yes	
122	GSVIVG01013453001	331334..332034	VVC2962A	SEQ	CTACTGGCATCCTCTTTTTC	GCAATAGTTTGAACCAGGAC	yes	
122	GSVIVG01013453001	337542..337829	VVC2961A	SEQ	TCATTTCAATACCCAACCTC	CATTCCAATCTTCTGCACT	yes <sup>5</sup>	
122		341159..341247	VV05802A	SEQ	ATGGTAAAGTGCGTTGTC	CAACTCCTTCTATCTGCTT	yes	
122	GSVIVG01013452001	352614..352996	VV05801A	SEQ	TGAGGTTCGTGTCTAATC	TTTGAATGAGTTGGTCTCT	yes	
122	GSVIVG01013451001	372496..372799	VV05800A	SEQ	TTGATGTGGAAATGAGG	AAATGTGTGTATGCTTCTTG	yes	yes +R
122	GSVIVG01013450001	382744..382940	VVC2956A	SEQ	GGAAGCCAAGGAGTAATTGT	TGCTAAGTGGTCAAGAAGGT	yes	yes +R
122	GSVIVG01013449001	388123..388423	VVIN03	SSR	ACTCAAATTTGTTCCCTACTCT	TGAGAAAACTGAACCAGACTA		yes
122	GSVIVG01013449001	389679..390333	VVC2955A	SEQ	CCAATGTGTTTCTCTTCTTC	GGCAAGTCACTTAGTTTTC	yes	
122	GSVIVG01013448001	394283..394912	VVC2954A	SEQ	CTGATCTTCACAACAGCAAG	GAAGCAGCAAGGGTATTTTC	yes	
122	GSVIVG01013446001	399382..399793	VVC2953A	SEQ	ATCTCTTGAACAACGCTCTC	GGCACAATTTGACCTTATTTT	yes	yes +R
122	GSVIVG01013445001	412327..412964	VVC2951A	SEQ	AGCACTTGGCTTCCATCT	ACTGCCTTCTCTCTTTCT	yes	
122	GSVIVG01013443001	420841..421214	VVC2950A	SEQ	CTCGTTTCCGATTACCT	CAATTCTTTCAACATTTTATCC	yes	
122	–	423185..423271	1038A12F	SNP	AGACTGTAACAACCATAATTCAAAAATTTTACTTCTTGGCCCTCTGGAGCCAC ACTTTTCTGGCCTTTGTCTTCCACGCTTCCCAT[C/T]GGAACGGCTTCCCGACTT AACTGGAAGTGTACGGGAAGAGACAGAGCTAGACACATGATGCCAGGAACC AAAGCAGGGAACCCAACAGCA			yes

12X Genome Assembly			Fragment id	Marker type	Primer pairs <sup>s</sup>		Set of genotypes sequenced or genotyped	
Scaffold n°	Gene id	Position			Forward	Reverse	SnpGrapMap*	CHA-UB*
122	GSVIVG01013440001	427151..427764	VVC2949A	SEQ	GTCAGGGTAGAATTGGTGA	GTCAGGGTAGAATTGGTGA	no	no
122	GSVIVG01013438001	429725..431077	VV05799A	SEQ	GTATGATCGGATGGATAATTG	ACAAGAAAAACAAGGAAAGCA	yes <sup>s</sup>	yes +R
122		436989..437657	VVC2947A	SEQ	AACGGTCTCGACTTAGAACA	GTTATTGACATCTCCCCAAA	yes	
122	GSVIVG01013440001	444180..444719	VVC2946A	SEQ	TGAGAAGCTGGAGAAGGTAG	CGGTCCATCAAGAGATTGT	yes	
122	GSVIVG01013439001	467685..469298	VVC2945A	SEQ	CTGGAATTATGCTCTCTCTCA	TTGTAAGTGCACCAATAACC	no	no
122	GSVIVG01013438001	475406..476081	VVC2944A	SEQ	CATCAAGCCACTCAACAAC	ATAGATTGGAACACCAGCAC	yes	
122	GSVIVG01013436001	492294..492676	VVC2943A	SEQ	AACTGCATGGGAGGTCTT	CTTATTAGTCTGTTCTTGTTCC	yes	
122	GSVIVG01013435001	494374..494464	VVS54	SSR	TCACAAATTGCAAAGGATGC	CGCAATTATGAAAGCTGAGAAG		yes
122	GSVIVG01013435001	497378..497760	VVC2942A	SEQ	GATCGCATCCTTCATCTATC	TGAATACTTTGCAGCCTTG	yes	yes +R
122		497723..498045	IN0954	CAPS	GCGGATGCCACTCTTGAA	GATTAGGAGCATCGTCGTCC		yes
122	GSVIVG01013434001	505166..505809	VVC2941A	SEQ	CCGCTAGGGTTACAGTTTAG	GCCTTGGTTATTTTCTTTGA	yes <sup>s</sup>	
122	GSVIVG01013434001	506509..507205	VVC2940A	SEQ	GGGTAAAGTAATGGACGAGA	CCTTGTAGCTTTTGATGGAG	yes	
122	GSVIVG01013433001	510613..510723	VV05798A	SEQ	AGAGAGTGTGCGTGACTG	ATAAATGATAGGAGAAGCAA	yes	yes
122	GSVIVG01013430001	525649..526210	VVC2935A	SEQ	ATCACCGATAAGAACACTGG	GGGTACATAAATCCTTGCTG	yes	
122	GSVIVG01013427001	532253..533642	VV05797A	SEQ	CTTCAGTTGGGATTGTTC	AACCTTCTTCCAGTCATC	no	no
122	GSVIVG01013427001	537389..537613	VVC2933A	SEQ	ATTGTTCTTAACGCACTGG	TCTGAAGCCAATACTCAATGT	yes	
122	GSVIVG01013426001	542891..544684	VVC2932A	SEQ	TGCCATCTCTAATGTGACTCT	CATTCTGGTGGATTACTGT	no	no
122	GSVIVG01013425001	549494..550104	VV05796A	SEQ	TGCTTCTTATGGTATGTCAG	CATTTGCTTCACCTATTGT	yes	yes +R
122	GSVIVG01013423001	565511..566165	VVC2929A	SEQ	CAAGGGGTTGATTTCTCTC	GCACTCACTTCTCCAGCTAC	yes	
122	GSVIVG01013421001	581021..581630	VVC2927A	SEQ	TGGAGAAAGTGTTGGATGTT	TCATTGCTTCCATCTTCATT	yes	
122	GSVIVG01013419001	590575..590818	VV05795A	SEQ	CAAAGATGGGAGAAAGAG	ACCACCGCTACAATAATC	yes	
122	GSVIVG01013417001	600097..600661	VVC2921A	SEQ	TGTTAGGAAGAAAGAAACCA	AAATAACTCGAACCCATTCA	yes	
122	GSVIVG01013416001	604441..605002	VVC2920A	SEQ	ACGAATCCATGAAAGAATGT	GAACAGAAGAATCCAGTCCA	yes	
122	GSVIVG01013414001	605645..605876	VVS55	SSR	TCCAAGAAGCATTATATGCAA	TGCCCATTAGTTTTTAAATGG		yes <sup>ss</sup>
122	GSVIVG01013414001	615081..615296	VV05793A	SEQ	AGTAAAGGAAGGATGGTTG	ATGGACACTACCGAACAG	yes	yes
122	GSVIVG01013413001	622356..622533	VVC2918A	SEQ	GAAAAAGGATGTGGGAATAA	CTGAATACGGAATGGAACAC	yes	
122	GSVIVG01013411001	633807..634244	VVC2916A	SEQ	ACTTAAAAACCCACTTGCTG	CGAACATCTTGGTCTCTCTG	yes	
122	GSVIVG01013410001	638081..638634	VV05791A	SEQ	TCTGTTTATTTACGGGATG	AATGTTGATTATGCTCTGG	yes	
122	GSVIVG01013409001	645294..645441	VVC2914A	SEQ	CATTACAGATGGGATGTT	ACTCTCGACTTGACACTGCT	yes	
122	GSVIVG01013408001	650634..651237	VV05790A	SEQ	TAGGATTCAATTCTCTACTCT	ATGGAACACTGCTAAAAA	yes	
122	GSVIVG01013407001	654932..655330	VVC2912A	SEQ	CTTTGGGGGAACATTCTTT	GGTACTGATTGGACTGCTTC	yes	
122	GSVIVG01013406001	668381..668534	VV05788A	SEQ	GACTACACAACAATGCGTC	TAATACCCACAACCACTTT	yes	yes
122	GSVIVG01013404001	670015..670200	VVS56	SSR	GTGCTTTTAGCACCAAAGAGT	TGGGACAAGATTGTAGCCAAA		yes
122	GSVIVG01013404001	676149..676503	VVC2898A	SEQ	TCAGAAGCTCATTGTGAAAG	AGTACAAGTGGCAACACACA	yes**	
122	GSVIVG01013404001	676150..676250	VVC2910A	SEQ	CGGTATCTTGGTAGTTTGG	CATAGATGAATTTGGGCAGT	yes	
122	GSVIVG01013403001	682572..683107	VVC2897A	SEQ	ACGCCCTACTTACCTTCTTC	ACCATCTCATCAAGCATTTTC	yes	
122	GSVIVG01013403001	682588..683107	VVC2909A	SEQ	ATTACTCCCATGAGGTTTCC	ACCATATCATCTCCCATGAC	yes**	
122	GSVIVG01013402001	686722..687437	VV05786A	SEQ	TTCATCTATCTATTCTTCGTATC	CTTCTTCTTTTGGGACT	yes	
122	GSVIVG01013401001	694377..694978	VVC2895A	SEQ	GGTCACAAGATCAGTAGCAG	TCATCCAACCATGTCAAAC	yes**	
122	GSVIVG01013401001	694378..695079	VVC2907A	SEQ	TCACTACTTCTCCTGCCACT	TTATTCGGCTCAAAACAATC	yes	
122	GSVIVG01013400001	702907..703637	VV05785A	SEQ	GGCTTCTGTCAATTGTTGT	AATAGTCCCACCCTGATT	yes	yes
122	GSVIVG01013399001	708961..709303	VVC2893A	SEQ	TACAACCTCACAGCTCAACC	CCTCATTTCCATCTCTCT	yes**	
122	GSVIVG01013399001	708973..709576	VVC2905A	SEQ	ATTCTCTCTCCGTTTACC	CTCGTCTCCATGTACTGTT	yes	
122	GSVIVG01013397001	729337..729889	VVC2891A	SEQ	CCATAGAAAACCAACCAAGA	TCTCTAACTCCTTGCAGACC	yes	
122	GSVIVG01013397001	729373..730002	VVC2903A	SEQ	TTGGAAGAGGAATAACAGGA	GCTCCACAACCTTTCTAACA	yes**	
122	GSVIVG01013395001	733597..734350	VV05783A	SEQ	TTGAGAAGGCTGATAAAA	GGTGAGTCCAACAAGAAC	yes	
122	GSVIVG01013394001	742593..742888	VVC2901A	SEQ	AGGGAATGGTGAAGAAGTATC	CAGTTGTTTCATAGTGAGCTTG	yes	

12X Genome Assembly			Fragment id	Marker type	Primer pairs <sup>§</sup>		Set of genotypes sequenced or genotyped	
Scaffold n°	Gene id	Position			Forward	Reverse	SnpGrapMap*	CHA-UB*
122	GSVIVG01013393001	746740..747273	VVC2885A	SEQ	CGCCTTAAAAATAGCAATAGG	AACCTACCTTTGGTGTCAACT	yes	
122	GSVIVG01013393001	746812..747175	VVC2900A	SEQ	ACCTCAAACCTTCACTCTGCT	CTTTGGAGAAAACCAATGTTC	yes**	
122	GSVIVG01013392001	758261..758760	VVC2899A	SEQ	AAGCGAGAATTAACACCAGA	AGCCTCCAACCTCAAAAATAG	yes**	
122	GSVIVG01013392001	758426..758892	VVC2884A	SEQ	AACTTTCAGATTTTGGGTTG	CCATGTGGAGGAGTACGA	yes	
122	GSVIVG01013389001	776756..777088	VV05782A	SEQ	GAGACAAGGCTGTGAATG	TGACAATGGATACTTCTGG	yes	yes
122	GSVIVG01013401001	792247..793997	VVC2894A	SEQ	TGAAGGTCAGAGAAAGAAGG	AGTCATCGACAACCTGAAAG	no	
122	GSVIVG01013400001	804498..804634	VVS57	SSR	AGGCATTTCCGATACAGCTC	TTCAACGTATACAAGCGACTTC		yes
122	GSVIVG01013385001	808955..809338	VVC2892A	SEQ	ATGGAAGACGATTCTTGACC	TGCAAGGCAAGTAAAAACA	yes	
122	GSVIVG01013382001	818661..819292	VV05781A	SEQ	ACCCACTCTACCCCTCTTC	ATTTTGTCTCTGTGTCAGT	yes	yes
122	GSVIVG01013381001	825195..825871	VV05780A	SEQ	CACGAACGATACAAGAAC	AACCAGAAGTAGTAGGTGG	yes	
122	GSVIVG01013394001	837719..838815	VVC2887A	SEQ	GATCTTTGAATGGGCTTATCT	ACCGTTCTCTTCTCATCAA	no	
122	GSVIVT01013373001	877751..878077	VVCS1H085H20R1-1	SSR	TTCTTTTGAAATTGCATCCT	TTAAGTTGGCAAAAGAGGAA		yes
122	GSVIVG01013393001	895761..895846	1073P15R	SNP	TTGACTGCCATGAATGGTTGCATGATATTATATGCAAAAGATAGGATCAGCTTTG TTTCAAATTAACATCCCAGCATCGGACACTAAAATTTCAAAAAAATTTACAAC TACTCCAGATCTCAAGCCTCAATAATT[C/T]CGTACCTTTTGAAATTTCAAACAA TTCAAACACATTTTACACAAGAGTTAGAGAAATTTTAATTCAAAGTTTTTC			yes
122	GSVIVG01013373001	898379..898848	VV05779A	SEQ	GTGCGGTATTTATTGTCT	TTTCTATCTCCACTGCTTAC	yes	yes
122	GSVIVG01013392001	901775..901934	VVS58	SSR	GGGCTGAACATAAGTGAGCTG	TACGTAGGCGTGCAAATCG		yes
122	GSVIVG01013370001	909911..910626	VV05778A	SEQ	CTGTGGATTCTGCTTGTT	GCACTGACTTGAAC TGGA	yes	
122	GSVIVG01013367001	928463..929045	VV05777A	SEQ	CTATTGTGGTTCCTGGTG	GAAGTTGTTCTCCTTTGG	yes	yes +R
122	GSVIVG01013365001	940974..941683	VV05776A	SEQ	ATGGAATGAGTGAGAGCA	GTAATGGAAACCGTATGAA	yes	
122	GSVIVG01013381001	948267..948387	VMC2A3	SSR	ATTGAAACTCCGGAAGCTTAGG	CTTCGTGTAGAAAGCTTCACAGGT		yes
122	GSVIVG01013363001	949921..950653	VV05775A	SEQ	TATGTATGCTTGGGAAGA	ATAAAGAAGTGGAACCC	yes	yes
122	GSVIVG01013370001	1009539..1010696	VVC2869A	SEQ	AGGCAATCCTCAAAAATAAG	ATGTTGATGTGGAAACCAT		yes
122	GSVIVG01013367001	1054896..1056021	VVC2865A	SEQ	GAGCATGAGGCAATCCTT	TCTGTCTCCTTTGATCTTC		yes
1	GSVIVG01013365001	1084916..1085337	VVC15574A	SEQ	GAGCATGAGGCAATCCTT	GAGCTATCTGTCTCCTTTG		yes +R
1	GSVIVG01008672001	1098027..1099246	VVC15572A	SEQ	TAGTCCCAACCTACAACGAG	CCTGAAATCAAGTCTCCATC		yes +R
1	GSVIVG01008673001	1104978..1106307	VVC15571A	SEQ	TTAACGGATGCTTCTCCTAC	GTCGGCATTTTTTCATAACTC		yes +R
1	—	1226489..1226647	VVCS1H066N21R1-1	SSR	CCCCTCTTCGTCTCCAACAA	GCAGCCCTTCCCCATCAAT		yes
1		1297892..1298020	C011	SSR	GCTGGAGCTGCTATCAGATC	CAGGAGGAGGTGATTGTTAC		yes
1		1452854..1453153	VVCS1H085F05F1-1	SSR	CCGGTTATTAATTTACCTTC	TAACGGTTTCGAATCAAATCA		yes
1	GSVIVG01008868001	2912753..2913088	VVIB31	SSR	GGTTGGTACCAATGAAATCAAT	ACGTTCTACAGTATTCTCAA		yes
1		3505999..3506060	VVIV16	SSR	CGTTTCTTTGTCATCCATGTGA	AAAGCCTACGTTAGACAAAAC		yes

\* "yes" if the sequencing was done only in the forward orientation, "yes+R" if the sequencing was done in both orientations and "no" if the amplification failed

<sup>§</sup> Two gene models in the 8X were fused in one gene model in the 12X

\*\* Gene models from a false duplication on the 8X genome sequence version, suppressed on the 12X version.

<sup>ss</sup> Discarded because multilocus

<sup>§</sup> For SNP markers, primers used for amplification are indicated in bold and the extension primer is underlined.



Fragment code	Chromosome	Position (12X assembly)	Gene id (12X assembly)	Primer pairs		Sequence length	exon sequence (bp)	intron/UTR sequence (bp)	Wild <i>Vitis vinifera</i>				Cultivated <i>Vitis vinifera</i>				
				Forward	Reverse				Number of genotypes	Number of segregating sites	Number of haplotypes	$\pi$	Number of genotypes	Number of segregating sites	Number of haplotypes	$\pi$	Tajima's D
VV04784A	13	7789388..7790047	GSVIVG01034690001	CTTCAGTGTCTGTCCATT	GAGTTGTCTATCCCATCC	660	0	660	7	5	3	0.00225	27	12	7	0.00318	-0.67757
VV04811A	1	19473823..19474481	GSVIVG01010320001	CTAATCGTGTGCTTCTACTAC	CTTCTTTGATGACTTGGA	1005	832	173	7	3	2	0.002	27	7	7	0.00254	0.23693
VV04929A	10	3197524..3198136	GSVIVG01021265001	TCTTCTGGTGCTGTAG	GTCCGCTATTCTCTTCTT	613	547	66	6	2	3	0.00114	27	4	5	0.00185	0.65711
VV05031A	16	21021110..21021739	GSVIVG01038685001	AAAGTTCAGCCACCTACA	ATCATCCACATCAAAAGG	630	58	572	7	2	3	0.00072	27	5	5	0.0014	-0.541
VV05338D	2	7633910..7634509	GSVIVG01013145001	GACTGTTCACAGGTTTCT	GGTCTTTGGGATTACCTC	600	428	172	6	3	4	0.00084	28	7	7	0.00134	-1.24735
VV05408A	8	18638556..18639137	GSVIVG01033690001	TTCTTGGATTCTGGTAAAA	GGGTTCCTGTGTGTATTG	582	375	207	7	2	3	0.00102	28	3	4	0.00059	-0.76854
VV05469A	2	3894757..3895475	GSVIVG01019858001	TACTCTACTCCCGTTCCA	CACCTCTCTCATCGTCATC	719	83	636	7	1	2	0.0002	27	6	4	0.0013	-0.72142
VV05490A	2	18137554..18138188	GSVIVG01036891001	GAGAAGAATGAGAACCTGA	ATAAGGACTGAGAACATACC	635	569	66	7	4	4	0.00241	27	11	7	0.00274	-0.79865
VV05523A	3	7202635..7203248	GSVIVG01037845001	AACCTTGATTGGAGAGAAG	AGTGAITTTGGCTGTITT	614	515	99	7	3	4	0.0019	27	11	10	0.00419	0.29997
VV05852A	19	10336646..10337265	GSVIVG01014835001	TATTCTCAGCCAAAATCC	ATAAACAACTCTATGCCTCTT	620	620	0	7	9	5	0.00401	26	14	14	0.00594	0.56712
VV05889A	7	20982766..20983271	GSVIVG01000277001	GGTAAGAGTGATGCTGTGA	CATAGAAGGGAAACCAGA	506	301	205	7	5	3	0.00399	28	10	7	0.00405	0.14672
VV06195A	16	16096142..16096767	GSVIVG01010605001	TGGGGTATGTTTGTAG	CTCAGTTAGGATAGTGTATC	626	477	149	7	2	3	0.00237	28	8	8	0.00202	-0.90669
VV06254A	10	5454489..5454897	GSVIVG01021445001	TATCGCTCAAAATGTCTATC	TATGTCCTTGTCTTCTTG	409	127	282	7	7	5	0.01506	28	9	8	0.0132	0.23243
VV06623D	5	11589731..11590350	GSVIVG01029057001	TCGCTTATTTCAACGGTC	GAGAATCAACAAAAGTCCC	620	620	0	6	3	2	0.00257	27	16	9	0.0047	-0.81093
VV06625A	Random	29700889..29701646	GSVIVG01007007001	AAACGCACCTATCCTTTG	TCTTGACCTCCATTACC	758	37	721	7	2	3	0.00067	28	7	7	0.00212	0.14259
VV06645A	Random	7646308..7647093	GSVIVG01003182001	TGTTTCGTTGATTCTCTTT	TAGGGTTTCTATGGCTCT	786	786	0	7	5	4	0.0021	27	9	12	0.00351	1.08858
VVC2943A	18	492294..492676	GSVIVG01013436001	AACTGCATGGGAGGTCTT	CTTATTAGTCTGTCTTGTTTCC	383	117	266	6	2	2	0.0025	28	6	6	0.00242	-0.69859
VVC6002B	4	23597915..23598431	GSVIVG0102644001	CATGTTAAGCTCCACCAGAG	AGAGGAGGAAGAGGAGATGA	517	456	61	6	6	5	0.0051	23	10	11	0.00349	-0.59919

**Additional file 4: Table S4.** qPCR validation of homozygotie in the *Flb* region in Ugni Blanc mutant. Estimation of the initial number of DNA quantity of the FL gene and of the HMGCoA gene family in Pinot Noir (PN777), Chardonnay (CHA), Ugni Blanc mutant (UBM) and Ugni Blanc (UB), before and after normalization by the result obtained for the HMGCoA genes.

	<i>FL</i>		<i>HMGCoA</i>		<i>FL</i> normalized	
	Qty Mean	Qty StdDev	Qty Mean	Qty StdDev	Qty Mean	Qty StdDev
<b>PN777</b>	1,54	0,15	1,37	0,06	<b>1,13</b>	<b>0,17</b>
<b>CHA</b>	2,43	0,09	2,28	0,07	<b>1,07</b>	<b>0,08</b>
<b>UBM</b>	2,59	0,14	2,38	0,05	<b>1,09</b>	<b>0,08</b>
<b>UB</b>	1,79	0,03	1,47	0,05	<b>1,23</b>	<b>0,06</b>



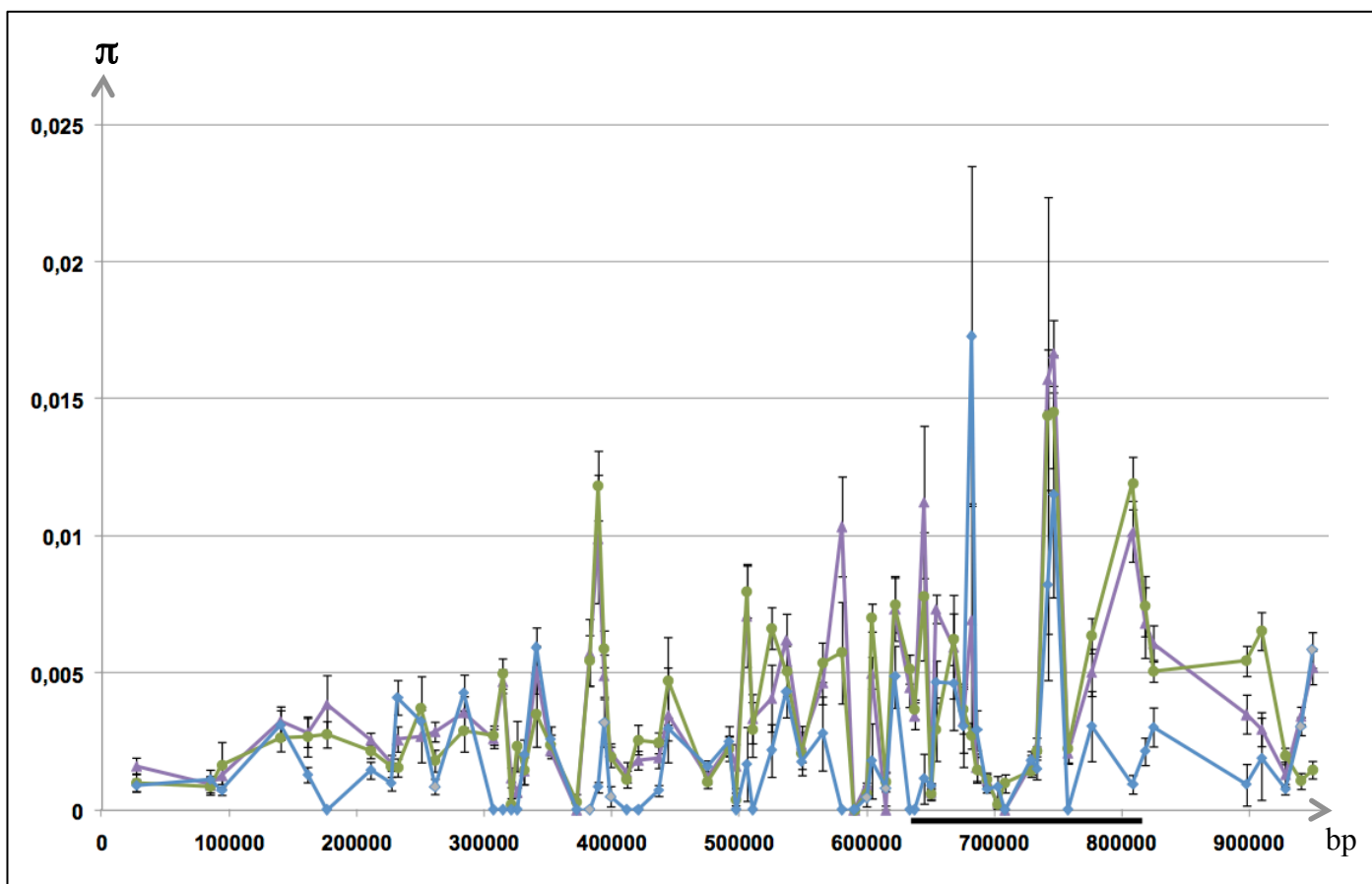


Fragment code	Sequence length	% exon sequence	% intron/UTR sequence	Wild <i>Vitis vinifera</i>							Whole set of cultivated <i>Vitis vinifera</i>										
				Number of genotypes	Number of segregating sites	Number of haplotypes	Haplotypic diversity (Hd)	Hd standard error	$\pi$	$\pi$ standard error	Number of genotypes	Number of segregating sites	Number of haplotypes	Haplotypic diversity (Hd)	Hd standard error	$\pi$	$\pi$ standard error	Tajima's D <sup>§</sup>	Number of genotypes	Number of segregating sites S	Number of haplotypes
VVC2951A	638	39	61	5	0	1	0	0	0	0	26	5	6	0.453	0.081	0.00114	0.00024	-0.82274	10	4	4
VVC2953A	412	15	85	5	1	2	0.2	0.154	0.00049	0.00037	26	4	6	0.662	0.045	0.00202	0.00023	-0.13562	11	3	4
VVC2954A	630	84	16	4	6	4	0.75	0.139	0.00317	0.00087	26	15	10	0.887	0.016	0.00546	0.00048	0.11035	10	13	7
VVC2955A	650	11	89	4	1	2	0.536	0.123	0.00082	0.00019	18	22	10	0.857	0.029	0.01118	0.0011	1.25921	10	21	6
VVC2956A	197	6	94	6	0	1	0	0	0	0	27	3	4	0.603	0.048	0.0053	0.00073	1.18743	11	3	3
VVC2962A	701	62	38	5	3	2	0.467	0.132	0.002	0.00056	24	8	6	0.558	0.076	0.00171	0.00037	-0.92156	10	4	3
VVC2963A	603	74	26	7	0	1	0	0	0	0	25	8	4	0.468	0.06	0.00152	0.00054	-1.32938	10	1	2
VVC2965A	499	0	100	6	0	1	0	0	0	0	24	8	7	0.812	0.024	0.0048	0.00031	0.90164	10	6	5
VVC2966A	557	0	100	6	0	1	0	0	0	0	25	5	5	0.767	0.027	0.00253	0.00023	0.63192	12	5	5
VVC2970A	699	48	52	6	5	4	0.773	0.069	0.00321	0.00055	27	13	8	0.65	0.061	0.00317	0.00074	-0.66434	11	12	6
VVC2972A	358	0	100	5	3	3	0.689	0.104	0.0041	0.00063	24	5	6	0.637	0.044	0.0023	0.00031	-0.65753	11	4	5
VVC2973A	610	0	100	7	2	3	0.538	0.115	0.00095	0.00025	24	3	4	0.645	0.034	0.0017	0.00007	1.10647	9	3	4
VVC2974A	674	17	83	6	2	2	0.485	0.106	0.00144	0.00031	26	9	8	0.722	0.042	0.00229	0.00017	-0.62078	10	5	5
VVC2975A	199	0	100	5	0	1	0	0	0	0	25	3	4	0.5	0.071	0.00322	0.00056	-0.9182	11	3	4
VVC2981A	702	0	100	6	1	1	0.485	0.106	0.00069	0.00015	26	13	5	0.425	0.08	0.0014	0.0005	-1.96104*	11	4	4
VVC2982A	745	10	90	6	2	2	0.409	0.133	0.0011	0.00036	25	6	5	0.464	0.08	0.00099	0.00022	-1.13712	11	4	4
VVC2985A	494	33	67	7	1	2	0.44	0.112	0.00089	0.00023	27	4	5	0.511	0.067	0.00127	0.00021	-0.64153	11	3	4
Total/Average	34355	38.4	61.6	5.9	2.8	2.2	0.37	0.093	0.0020	0.0006	24.8	8	5.8	0.603	0,08	0,0041	0,0004		10,5	6,2	4,3

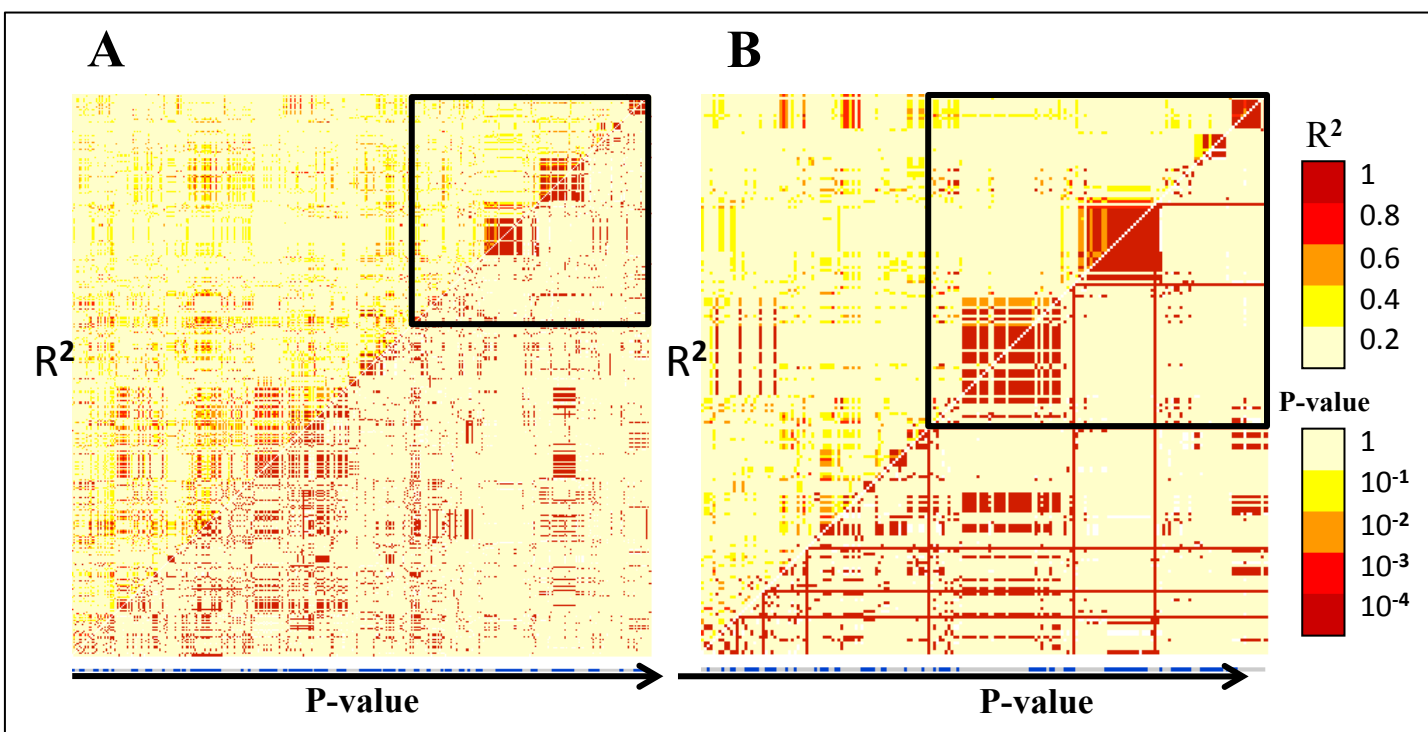
<sup>§</sup> \* for 0.01<p<0.05 And \*\* for 0.001<p<0.01



Wine cultivars					Table cultivars								Number of INDEL			Number of segregating SNP sites									
Haplotypic diversity (Hd)	Hd standard error	$\pi$	$\pi$ standard error	Tajima's D <sup>§</sup>	Number of genotypes	Number of segregating sites S	Number of haplotypes	Haplotypic diversity (Hd)	Hd standard error	$\pi$	$\pi$ standard error	Tajima's D <sup>§</sup>	Wild	Domestic ated	Shared Wild/Domestic ated	Wild	Domestic ated	Wine	Table	Shared Wild/Domestic ated	Shared Wine/Table	Shared Wild/Wine	Shared Wild/Table		
0.489	0.117	0.00137	0.00037	-0.65164	14	4	5	0.484	0.109	0.00112	0.00031	-0.79152	0	0	0	0	5	4	4	0	3	0	0		
0.697	0.058	0.00212	0.0003	0.02395	13	3	5	0.628	0.087	0.00193	0.00037	0.02395	0	2	0	1	4	3	3	1	2	0	1		
0.874	0.034	0.00489	0.00077	-0.57946	14	14	7	0.857	0.03	0.00587	0.00067	0.09734	1	1	1	6	15	13	14	5	12	3	4		
0.848	0.074	0.00988	0.00234	-0.3357	12	22	9	0.877	0.036	0.01183	0.00127	1.13124	1	1	1	1	22	21	22	1	21	1	1		
0.593	0.085	0.00574	0.00122	-0.9646	14	3	4	0.659	0.052	0.00545	0.00092	0.94396	0	0	0	0	3	3	3	0	3	0	0		
0.426	0.122	0.00142	0.00049	-0.34083	13	7	5	0.566	0.086	0.00144	0.00051	-1.37182	0	1	0	3	8	4	7	3	3	3	2		
0.395	0.101	0.00065	0.00017	0.72261	13	8	4	0.566	0.068	0.00232	0.00092	-1.04449	0	0	0	0	8	1	8	0	1	0	0		
0.789	0.049	0.00469	0.00043	1.2265	13	8	7	0.828	0.042	0.00497	0.00055	0.5752	0	4	0	0	8	6	8	0	6	0	0		
0.797	0.039	0.0026	0.00035	0.17034	12	5	5	0.757	0.058	0.00272	0.00035	0.37995	0	0	0	0	5	5	5	0	5	0	0		
0.667	0.092	0.00269	0.00096	-1.49944	14	11	6	0.579	0.104	0.00372	0.00115	-0.26146	1	1	1	5	13	12	11	5	10	5	3		
0.688	0.078	0.00258	0.00046	-0.44841	12	2	3	0.507	0.093	0.00154	0.00033	0.06219	0	0	0	3	5	4	2	3	2	3	2		
0.634	0.093	0.00174	0.00026	0.59324	13	2	3	0.615	0.063	0.00157	0.00015	1.78129	0	0	0	2	3	3	2	1	2	1	1		
0.753	0.074	0.00255	0.00026	0.66532	14	7	6	0.677	0.057	0.00213	0.00025	-0.60517	0	0	0	2	9	5	7	2	3	2	2		
0.455	0.115	0.00387	0.00104	-0.1647	12	2	3	0.518	0.073	0.00275	0.00048	0.04732	0	0	0	0	3	3	2	0	2	0	0		
0.515	0.111	0.00126	0.00032	-0.53704	14	13	5	0.384	0.113	0.00161	0.00086	-2.22066**	0	0	0	1	13	4	13	1	4	1	1		
0.468	0.122	0.00093	0.0003	-1.03561	12	5	4	0.373	0.12	0.00087	0.00032	-1.50208	0	0	0	2	6	4	5	2	3	2	2		
0.636	0.08	0.0016	0.0003	-0.09667	14	3	4	0.373	0.107	0.001	0.00032	-0.8683	1	1	1	1	4	3	3	1	2	1	1		
0,60	0,07	0,0036	0,0007		12,5	6,6	4,7	0,57	0,08	0,0035	0,0007		10	33	10	193	554	430	453	187	335	174	131		



**Additional file 6: supplemental figure S1.** Nucleotide diversity in the cultivated (table and wine) and wild compartment along the *Flb* region. Nucleotide diversity in the table grapes (green line), the wine grapes (purple line) and the wild grapes (blue line) along the *Flb* region.



**Additional file 7: supplemental figure S2.** Linkage disequilibrium along the *Flb* region in cultivated and wild compartment. LD plots on  $R^2$  values (above the diagonal) and associated P-value (below the diagonal) along the entire *Flb* region in cultivated (A) and wild compartment (B). The gene fragments re-sequenced are represented by alternate grey and blue boxes, which size is proportional to the number of polymorphic SNP used in the LD estimation. The black arrow represents the orientation of the region from the telomere (on the left) to the centromere.

**Additional file 8: supplemental Table S6.** List of gene fragments and their associated number of SNPs used for the estimation of LD in the *Flb* region, in cultivated and wild *V. vinifera* pools.

gene name (12X)	primer name	Number of SNP marker used for LD estimation	
		cultivated pool	wild pool
GSVIVT00013479001	VVC2985A	3	1
GSVIVT00013476001	VVC2982A	3	2
GSVIVT00013475001 <sup>s</sup>	VVC2981A	4	1
GSVIVT00013472001	VV05807A	6	4
GSVIVT00013471001	VV05806A	4	1
GSVIVG00013469001	VVC2975A	3	
GSVIVT00013468001	VVC2974A	4	2
GSVIVT00013467001	VVC2973A	2	2
GSVIVT00013466001	VVC2972A	2	3
GSVIVT00013462001	VVC2970A	12	5
GSVIVT00013461001	VV05805A	5	2
GSVIVT00013460001	VV05804A	5	4
GSVIVT00013458001	VVC2966A*	5	
GSVIVT00013456001	VVC2965A	7	
GSVIVT00013455001	VV05803A	2	
GSVIVT00013454001	VVC2963A	6	
GSVIVT00013453001	VVC2962A	5	3
	VV05802A	1	1
GSVIVT00013452001	VV05801A	3	2
GSVIVT00013450001	VVC2956A	3	
GSVIVT00013449001	VVC2955A	22	1
GSVIVT00013448001	VVC2954A	14	6
GSVIVT00013446001	VVC2953A	2	1
GSVIVT00013445001	VVC2951A	3	
GSVIVT00013443001	VVC2950A	2	
	VVC2947A	6	1
GSVIVT00013440001 <sup>s</sup>	VVC2946A	13	3
GSVIVT00013438001	VVC2944A	4	2
GSVIVT00013436001	VVC2943A	3	2
GSVIVT00013435001	VVC2942A	1	
GSVIVT00013434001	VVC2940A	17	7
GSVIVT00013433001	VV05798A	1	
GSVIVT00013430001	VVC2935A	10	5
GSVIVT00013427001	VVC2933A	6	2
GSVIVT00013425001	VV05796A	7	2
GSVIVT00013423001	VVC2929A	9	6
GSVIVT00013421001	VVC2927A	16	
GSVIVT00013417001	VVC2921A	1	1
GSVIVT00013416001	VVC2920A	9	5
GSVIVT00013414001	VV05793A	2	1
GSVIVT00013413001	VVC2918A	5	3
GSVIVG00013411001	VVC2916A	6	
GSVIVT00013410001 <sup>s</sup>	VV05791A	4	
GSVIVT00013409001	VVC2914A	5	1
GSVIVT00013408001	VV05790A	3	1
GSVIVT00013407001	VVC2912A	7	4
GSVIVG00013406001	VV05788A	2	2
GSVIVT00013404001	VVC2910A	1	1

gene name (12X)	primer name	Number of SNP marker used for LD estimation	
		cultivated pool	wild pool
GSVIVT00013403001 <sup>\$</sup>	VVC2897A	5	24
GSVIVT00013402001	VV05786A*	5	6
GSVIVT00013401001	VVC2907A	4	1
GSVIVT00013400001 <sup>\$</sup>	VV05785A	2	2
GSVIVT00013399001	VVC2905A	2	
GSVIVT00013397001	VVC2891A	4	3
GSVIVT00013395001	VV05783A	5	4
GSVIVT00013394001 <sup>\$</sup>	VVC2901A	9	10
GSVIVT00013393001 <sup>\$</sup>	VVC2885A	19	15
GSVIVT00013392001	VVC2884A	6	
GSVIVT00013389001	VV05782A	5	4
GSVIVT00013385001 <sup>\$</sup>	VVC2892A	10	1
GSVIVT00013382001	VV05781A	16	3
GSVIVT00013381001	VV05780A	10	7
GSVIVT00013373001	VV05779A	8	3
GSVIVT00013370001	VV05778A	14	8
GSVIVT00013367001	VV05777A*	3	1
GSVIVT00013365001	VV05776A	7	4
GSVIVT00013363001	VV05775A*	11	9

<sup>\$</sup> Genes showing a significant deviation from neutrality of Tajima's D parameters

\* Genes with SNP significantly associated with berry weight variation



**Annexe 4:** Figures et tableaux des résultats supplémentaires réalisés sur la région *Fleshless berry* (Chapitre 5)

**Annexe 4-A: Tableau:** Liste des SNP provenant de la région *Fleshless berry* génotypés sur le set de 38 *V. vinifera* sauvages supplémentaires.

Nom Amorce	Gène assemblage 12X	Position du gène	Type de SNP	Position du SNP sur le chromosome	Position du SNP dans l'annotation	Numéro du SNP	
VV05796A 234	GSVIVG01013425001	545163 .. 551046	[C/T]	549871	Exon	1	
VV05796A 323	GSVIVG01013425001	545163 .. 551046	[A/C]	549782	Intron	2	
VVC2927A 289	GSVIVG01013421001	579117 .. 582064	[T/C]	581309	UTR	3	
VVC2927A 83	GSVIVG01013421001	579173 .. 581118	[A/G]	581103	Exon	4	Absence de signal
VVC2921A 350	GSVIVG01013417001	599936 .. 601322	[G/T]	600446	Exon	5	Eau contaminée
VVC2920A 318	GSVIVG01013416001	602662 .. 605100	[C/A]	604685	Intron	6	Triallélique ou multilocus
VV05793A 372	GSVIVG01013414001	614215 .. 618895	[A/C]	615453	Intron	7	
VVC2916A 224	GSVIVG01013411001	631280 .. 634380	[C/A]	634021	Intron	8	
VV05791A 131	GSVIVG01013410001	635624 .. 640934	[G/A]	638504	Exon	9	
VVC2914A 344	GSVIVG01013409001	642852 .. 645549	[T/A]	645095	Intron	10	
VVC2914A 398	GSVIVG01013409001	642852 .. 645549	[C/T]	645041	Intron		
VV05790A 350	GSVIVG01013408001	646526 .. 652067	[A/G]	650888	Intron	11	
VVC2912A 255	GSVIVG01013407001	654277 .. 659188	[C/T]	655076	Intron	12	
VVC2912A 470	GSVIVG01013407001	654277 .. 659188	[C/G]	654861	Intron	13	
VVC2910A 450	GSVIVG01013404001	673537 .. 679430	[A/T]	676601	Exon	14	
VVC2897A 55	GSVIVG01013403001	681331 .. 684269	[A/G]	682626	Intron	15	
VVC2897A 536	GSVIVG01013403001	681331 .. 684269	[T/*]	683108	Intron	16	
VV05786A 120	GSVIVG01013402001	686243 .. 690429	[T/C]	686840	Intron	17	
VV05786A 565	GSVIVG01013402001	686243 .. 690429	[G/A]	687286	Exon	18	
VV05786A 596	GSVIVG01013402001	686243 .. 690429	[C/A]	687316	Exon	19	
VVC2907A 524	GSVIVG01013401001	691256 .. 696034	[G/T]	694901	Exon	20	
VV05785A 173	GSVIVG01013400001	697127 .. 703692	[A/C]	703465	Exon	21	
VV05785A 316	GSVIVG01013400001	697127 .. 703692	[A/G]	703322	Intron	22	
VVC2905A 158	GSVIVG01013399001	708986 .. 712359	[C/G]	709130	Intron	23	
VVC2905A 504	GSVIVG01013399001	708986 .. 712359	[C/A]	709476	Intron	24	
VVC2891A 262	GSVIVG01013397001	728952 .. 730694	[G/A]	729598	Exon	25	
VVC2891A 436	GSVIVG01013397001	728952 .. 730694	[G/C]	729772	Exon		
VVC2891A 456	GSVIVG01013397001	728952 .. 730694	[C/T]	729792	Exon		
VV05783A 224	GSVIVG01013395001	731976 .. 735233	[G/A]	734127	Intron	26	
VV05783A 494	GSVIVG01013395001	731976 .. 735233	[C/A]	733857	Intron	27	
VV05783A 522	GSVIVG01013395001	731976 .. 735233	[G/T]	733829	Intron	28	
VV05783A 605	GSVIVG01013395001	731976 .. 735233	[A/T]	733746	Intron	29	
VVC2901A 238	GSVIVG01013394001	740677 .. 743630	[A/C]	742830	Intron	30	
VVC2901A 38	GSVIVG01013394001	740677 .. 743630	[A/G]	742630	Exon		
VVC2885A 437	GSVIVG01013393001	746021 .. 748732	[T/C]	747176	Intron	31	
VVC2885A 484	GSVIVG01013393001	746021 .. 748732	[T/C]	747223	Intron	32	
VVC2884A 407	GSVIVG01013392001	750175 .. 760004	[G/C]	758486	Exon	33	
VVC2892A 240	GSVIVG01013385001	807953 .. 809293	[C/A]	809103	Intron	34	
VVC2892A 134	GSVIVG01013385001	807953 .. 809293	[T/G]	809209	Intron	35	
VV05781A 408	GSVIVG01013382001	817583 .. 819451	[AC/**]	818885 – 818884	Intron	36	
VV05780A 147	GSVIVG01013381001	821694 .. 826998	[C/T]	825341	Exon	37	
VV05779A 123	GSVIVG01013373001	875996 .. 900898	[T/G]	898501	Intron	38	
VV05779A 76	GSVIVG01013373001	875996 .. 900898	[A/G]	898454	Intron	39	
VV05777A 226	GSVIVG01013367001	927547 .. 931993	[T/A]	928820	Intron	40	
VV05776A 70	GSVIVG01013365001	938209 .. 944601	[G/A]	941043	Exon	41	
VV05776A 490	GSVIVG01013365001	938209 .. 944601	[A/C]	941463	Intron	42	
VV05775A 455	GSVIVG01013363001	949292 .. 961102	[G/A]	950199	Exon	43	

**Annexe 4-B : Tableau:** Liste des 38 *V. vinifera* ssp *sylvestris* supplémentaires

Nom	Code Vassal	Pays d'origine
Lamb Bois Bourdet 1	0Mtp1826	France
Lambrusque Akchour S21 Ind 4	8500Mtp241	Maroc
Lambrusque Akchour S21 Ind 5	8500Mtp242	Maroc
Lambrusque Akchour S21 Ind 6	8500Mtp243	Maroc
Lambrusque Béni Hassan S9 Ind 6	8500Mtp232	Maroc
Lambrusque Béni Idder S11 Ind 11	8500Mtp239	Maroc
Lambrusque Bicharole	8500Mtp14	France
Lambrusque forêt P5	8500Mtp17	France
Lambrusque forêt P7	8500Mtp15	France
Lambrusque Grésigne 14	8500Mtp129	France
Lambrusque Grésigne 7	8500Mtp127	France
Lambrusque I	8500Mtp53	Maroc
Lambrusque l'Escale 1	8500Mtp21	France
Lambrusque l'Escale 2	8500Mtp24	France
Lambrusque l'Escale 3	8500Mtp22	France
Lambrusque Oued Smir S10 Ind 23	8500Mtp236	Maroc
Lambrusque PSL 1	8500Mtp106	France
Lambrusque PSL 11	8500Mtp108	France
Lambrusque PSL 12	8500Mtp109	France
Lambrusque PSL 14	8500Mtp111	France
Lambrusque PSL 15	8500Mtp112	France
Lambrusque PSL 5	8500Mtp117	France
Lambrusque PSL 7	8500Mtp120	France
Lambrusque Puivert 1	8500Mtp20	France
Lambrusque Puivert 2	8500Mtp19	France
Lambrusque Saint-Louis	8500Mtp139	France
Lambrusque Ul'any nad Zitavou A07	8500Mtp181	Slovaquie
Lambrusque Ul'any nad Zitavou A51	8500Mtp210	Slovaquie
Lambrusque Ul'any nad Zitavou A55	8500Mtp213	Slovaquie
Lambrusque Ul'any nad Zitavou A77	8500Mtp222	Slovaquie
Ouch 14	0Mtp1566	France
Ouch 14 ter(9)	0Mtp1568	France
Silvestris du ruisseau des singes - Blida	8500Mtp152	Algérie
Vigne sauvage	8500Mtp1	France
Vigne sauvage Suisse 1	8500Mtp132	Suisse
Vitis silvestris Ketsch 2-39 mâle	8500Mtp40	Allemagne
Vitis silvestris Ketsch 27 femelle	8500Mtp34	Allemagne
Vitis silvestris Oberlin 21 mâle	8500Mtp37	Allemagne

Annexe 4-C: Tableau. Couples d'amorces utilisés en q-RT-PCR pour des gènes candidats pour la variation du poids des baies dans la région *Fleshless berry*.

Gènes candidats

Identifiant du gène dans la version 12X du génome	Nom du fragment de gène	Nom du couple d'amorces	Séquence de l'amorce sens	Séquence de l'amorce antisens	Longueur de l'amplicon (ADNg/ADNc)	Efficacité de la PCR
GSVIVG01013402001	VV05786A	VV05786B*	TCTCTTCTCCTAGAGCCTGA	ATTCATTCTCCCTGTCTTCC	406/145	0.99
		VV05786C	TCTCTTCTCCTAGAGCCTGA	GGATTCATTCTCCCTGTCTT	408/147	0.92
		VV05786D	AAATGGGATGAAAAGATGG	CTCCCTGTAGTGATTAACTTG	123/123	0
GSVIVG01013403001	VVC2897A	VVC2897B	AAGTGCTCTGACGTTTCTTC	CTGTAACCCCATTCAATTGTC	104/104	0.99§
		VVC2897C*	ATGATATGGTCAAGGTGGTC	TTTGTCTTGCAGAAATCCTC	554/148	0.99
		VVC2897D	TCAAGTGCTCTGACGTTTCT	CTGTAACCCCATTCAATTGTC	106/106	0.98
GSVIVG01013400001	VV05785A	VV05785B	TCTCCAGACACTCTTCTCGT	CTTCTTATGCCCAGTCTTTG	64/64	0.93
		VV05785C*	TCTCCAGACACTCTTCTCGT	GGCTTCTTCTTTGGGTACT	135/135	0.99

Gènes de référence

Identifiant du gène dans la version 12X du génome	Nom du gène	Référence	Séquence de l'amorce sens	Séquence de l'amorce antisens	Longueur de l'amplicon (ADNg/ADNc)	Efficacité de la PCR
GSVIVG01025146001	<i>EF1-α</i>	Reid et al, 2007	GAAGTGGGTGCTTGATAGGC	AACCAAAATATCCGGAGTAAAAGA	150/150	0.99
GSVIVG01016276001	<i>Aquaporin</i>	Reid et al, 2014	TCCGCCAAGGACTATCATGAC	CGCAATCAGAGCCCTGTAGAA	90/90	0.99

\* Couple d'amorces sélectionné pour la q-RT-PCR

§ Problème de dissociation du produit PCR

## Résumé

---

La taille du fruit est un caractère agronomique important pour le rendement des espèces cultivées. Or, excepté chez la tomate, les mécanismes impliqués dans la variation de la taille des fruits charnus sont encore mal connus. L'objectif de ma thèse était d'identifier les principaux facteurs anatomiques impliqués dans la variation de la taille des baies, de réaliser des tests d'association génétique sur des gènes candidats et de rechercher des possibles traces de domestication liées à la variation de la taille des baies. Le nombre de cellules déjà déterminé dans l'ovaire avant l'anthèse et la capacité des cellules du mésocarpe à grossir après l'anthèse se sont révélés les principaux déterminants de la variation de la baie entre les variétés. Trois gènes potentiellement impliqués dans ce caractère ont été identifiés par des approches de génétique d'association. La MADS box *VvAGL11* est associée au nombre de pépins de la baie. *VvHMGR1*, une HMGR impliquée dans la synthèse des isoprénoides précurseurs de régulateurs de croissance, est associée à la longueur des baies. Enfin, *VvJMJ* un facteur de transcription de type jumonji est à la fois associé à la variation du poids et à la longueur des baies. La caractérisation du locus *Fleshless berry* (*Flb*) a également été poursuivie avec (i) l'analyse du polymorphisme de séquence de 69 fragments de gènes entre les vignes sauvages et cultivées, ce qui a permis l'identification d'un gène présentant une trace de sélection et une diversité nucléotidique plus élevée dans le compartiment sauvage que dans le cultivé. (ii) La mise en évidence d'une homozygotie complète d'une région de 1 Mb contenant le locus *Flb* chez certains cultivars comme l'Ugni Blanc, rendant impossible le clonage positionnel du gène dans la population F1. L'ensemble de ce travail a permis de mettre en évidence la complexité des mécanismes à l'origine de la variation de la taille de la baie chez la vigne. Nous avons montré la faisabilité des études de génétique d'association chez la vigne en utilisant des gènes candidats pour le poids de baies, et identifié les candidats les plus intéressants pour la sélection assistée par marqueurs et les études fonctionnelles ultérieures.

**Mots clés:** *Vitis vinifera* L., taille de la baie, nombre et volume des cellules de la baie, génétique d'association, traces de sélection, *Fleshless berry*.

## Abstract

---

Fruit size is an important trait in fruit crops. However, excepted in tomato, little is known about the genetic and molecular control of fleshy fruit size variation. The aim of my thesis was to identify the main anatomical factors underlying berry size variation, to carry out an association genetics based approach with candidate genes and to search for possible signatures of domestication linked to berry size variation. Cell divisions before anthesis and cell expansion after anthesis were found to be the major determinants of flesh weight variation between varieties in grapevine. Three genes were identified by association genetics approaches. The *VvAGL11* MADS box was found associated only with the seed number per berry. *VvHMGR1*, a putative HMGR involved in the synthesis of isoprenoid, a precursor of plant growth regulators, was found associated with berry length. Finally, *VvJMJ* encoding a putative transcription factor jumonji was associated both with berry weight and berry length variation. The characterization of the *Fleshless berry* locus (*Flb*) was also pursued with (i) the analysis of the sequence polymorphism of 69 regions between wild and cultivated grapevines, that led to the identification of one gene showing a trace of selection and a higher nucleotide diversity in the wild compartment than in the cultivated compartment. (ii) The discovery of a complete homozygosity of a 1 Mb region containing the *Flb* locus in some cultivars such as the Ugni Blanc, making impossible the fine mapping of the *Flb* locus in a F1 population. This study has highlighted the complexity of the mechanisms underlying the berry size variation in grapevines. We demonstrated the feasibility to screen in grapevine by association genetics candidate genes for berry weight and to select the most interesting ones for marker-assisted selection and for further functional studies.

**Key words:** *Vitis vinifera* L., berry cell number and volume, berry size, association genetics, signatures of selection, *Fleshless berry*.